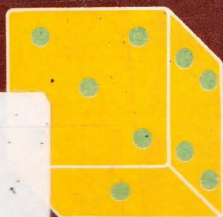
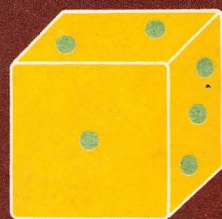
Introdução à Teoria da Probabilidade

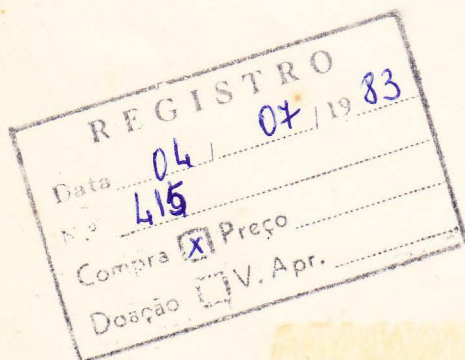
Hoel

Port

Stone

EDITORA INTERCIÊNCIA

Introdução à Teoria da Probabilidade

FUNDO EDITORIAL

Batschelet- Introdução à Matemática para Biocientistas
Bensaïd – A Consulta Médica
Bingham/Davies – Manual de Análise de Sistemas
Buecken – Vocabulário Técnico – Português, Inglês, Francês e Alemão
Coutinho – Jardim, Horta e Pomar
Dacorso – Elementos de Geometria Diferencial
Dawson/Wool – De Bits até If's – Uma Introdução ao Estudo dos Computadores e Fortran IV
Gomes/Helluy – Manual de Arquivo e Documentação
Harcourt e Laing (coordenadores) – Capital e Crescimento Econômico
Langridge – Classificação – Abordagem para Estudantes de Biblioteconomia
Leonhardt – Construções de Concreto – Vol. 1, Vol. 2 e Vol. 3
Lindgren – Temas de Planejamento
Mason/Mello e Souza – Métodos de Energia
McCullers/Van Daniker – Introdução à Contabilidade Financeira
McKinnon – A Moeda e o Capital no Desenvolvimento Econômico
Mitidieri – Problemas e Exercícios em Bioquímica
Motta Rezende – Materiais Usados em Eletrotécnica
Pemberton – Arranjo Físico Industrial e Movimentação de Materiais
Piedade – Introdução à Teoria da Classificação
Polya – A Arte de Resolver Problemas
Rego Monteiro – Tesouras de Telhado
Richardson – Economia Urbana
Silva Telles – Materiais para Equipamentos de Processo
Silva Telles/Paula Barros – Tabelas e Gráficos para Projetos de Tubulações – 2ª edição
Stephanes/Ferreira – Planejamento, Orçamento e Programação Financeira
Suszczynski – Os Recursos Minerais Reais e Potenciais do Brasil e sua Metalogenia
Swann – Técnicas de Aumento da Produtividade
Swingewood – O Mito da Cultura de Massa
Thomson – Teoria da Vibração com Aplicações
Wilmer/Pereira – Geometria para Desenho Industrial

Introdução à Teoria da Probabilidade

Paul G. Hoel **Sidney C. Port**
Charles J. Stone

Universidade da Califórnia — Los Angeles

TRADUÇÃO

Fernando Yassou Chiyoshi

EDITORA INTERCIÊNCIA

Programação Visual – Interciência Arte
Capa – Interciência Arte
Composição do texto – Interciência

CIP-Brasil. Catalogação-na-fonte
Sindicato Nacional dos Editores de Livros, RJ.

Hoel, Paul G.
H634i Introdução à teoria da probabilidade / Paul G. Hoel, Sidney C. Port, Charles J. Stone; tradução / de / Fernando Yassou Chiyoshi. – Rio de Janeiro: Interciência, 1978.

Tradução de: Introduction to probability theory
Bibliografia

1. Probabilidades I. Port, Sidney C.
II. Stone, Charles J. III. Título

CDD – 519
78-0330 CDU – 519

EDITORA INTERCIÊNCIA LTDA.
Rua Verna Magalhães, 66 - Tels.: 281-7495/263-5899
ZC-16 – 20000 – Rio de Janeiro – Brasil

PREFÁCIO

O propósito deste volume é servir como um texto, para curso de um trimestre ou um semestre, sobre teoria da probabilidade a nível júnior-senior. O material foi planejado para dar ao leitor preparação adequada, tanto para um curso de estatística como para estudos mais avançados de teoria da probabilidade e processos estocásticos. O pré-requisito para este volume é um curso de cálculo elementar que inclua integração múltipla.

Dedicamos esforços para apresentar somente os conceitos mais importantes da teoria da probabilidade. Tentamos explicar esses conceitos e indicar sua utilidade através de discussão, exemplos e exercícios. Alguns detalhes foram incluídos nos exemplos, de modo que se pode esperar, que o estudante os leia por conta própria, deixando assim ao instrutor mais tempo para cobrir as idéias essenciais e resolver um número considerável de exercícios em sala.

Há um grande número de exercícios ao final de cada capítulo, dispostos de acordo com a ordem em que o material relevante foi introduzido no texto. Alguns desses exercícios são de natureza rotineira, enquanto outros desenvolvem as idéias introduzidas no texto de forma um pouco mais profunda ou em direção um pouco diferente. Oferecemos sugestões para problemas mais difícies. Respostas, quando não são indicadas no próprio enunciado dos problemas, são fornecidas ao final do livro.

Embora a maior parte da matéria neste volume seja essencial para estudo mais avançado de probabilidade e estatística, algum material opcional foi incluído para dar maior flexibilidade. Essas seções opcionais são indicadas através de um asterisco. O material da Seção 6.2.2 é necessário somente para Seção 6.6; nenhuma dessas seções é necessária para este volume, mas ambas são necessárias em Introdução à Teoria Estatística. O material da Seção 6.7 é usado somente na demonstração do Teorema 1 do Capítulo 9 deste volume e Teorema 1 do Capítulo 5 de Introdução à Teoria Estatística. Os conteúdos dos capítulos 8 e 9 são opcionais; o Capítulo 9 não depende do Capítulo 8.

Desejamos agradecer a diversos colegas que leram o manuscrito original e fizeram sugestões, nos levando a um melhor resultado. Gostaríamos, também, de agradecer a Neill Weiss e Luis Gorostiza por terem resolvido e dado respostas a todos os exercícios e à Sra. Ruth Goldstein pelo excelente trabalho de datilografia.

ÍNDICE

ESPAÇOS DE PROBABILIDADE

A teoria da probabilidade é o ramo da matemática relacionado com fenômenos aleatórios (ou casuais). Muitas pessoas têm se dedicado ao seu estudo, devido ao seu interesse intrínseco, bem como às muitas aplicações bem sucedidas em muitas áreas das ciências físicas, biológicas e sociais, na engenharia e no mundo dos negócios.

Muitos fenômenos têm a propriedade de a sua observação, repetida sob um conjunto especificado de condições, conduzirem invariavelmente ao mesmo resultado. Por exemplo, se deixarmos cair uma bola, inicialmente em repouso, de uma altura de d pés através de vácuo ela atingirá o solo invariavelmente em $t = \sqrt{2d/g}$ segundos, onde $g = 32$ pés/s^2 é a aceleração constante devido à gravidade. Existem outros fenômenos cuja observação, repetida sob um conjunto especificado de condições, não conduz sempre ao mesmo resultado. Um exemplo familiar deste tipo de fenômeno é o lançamento de uma moeda. Se uma moeda é lançada 1000 vezes, as ocorrências de caras e coroas se alternam de uma forma aparentemente irregular e imprevisível. São fenômenos desse tipo que consideramos como sendo aleatórios e que constituem o objeto de nossa investigação.

À primeira vista pode parecer impossível fazer qualquer afirmação válida sobre tais fenômenos aleatórios, porém este não é o caso. A experiência mostra que muitos fenômenos não-determinísticos exibem uma *regularidade estatística*, que os torna passíveis de estudo. Isto pode ser ilustrado considerando novamente o lançamento de uma moeda. Para qualquer lançamento individual da moeda não podemos fazer nenhuma previsão não-trivial, mas as observações mostram que para um grande número de lançamentos a proporção de caras parece oscilar em torno de algum número fixo p entre 0 e 1 (sendo p muito próximo de 1/2 se a moeda é razoavelmente balanceada). Os resultados se comportam como se a proporção de caras em n lances, convergisse para p, ao fazer n tender a infinito. Pensamos nesta proporção limite p como a "probabilidade" que a moeda caia, em um único lançamento, com a cara voltada para cima.

De uma forma mais geral, a afirmação que um certo resultado experimental tem probabilidade p, pode ser interpretada, como significando que, se o experimento é repetido um grande número de vezes, aquele resultado seria observado

"cerca de" $100p$ por cento das vezes. Esta interpretação é chamada interpretação de freqüência relativa. Ela é muito natural em diversas aplicações da teoria da probabilidade aos problemas do mundo real, especialmente aqueles que envolvem as ciências físicas, porém freqüentemente parece ser bastante artificial. Por exemplo, como poderíamos dar uma interpretação de freqüência relativa para a probabilidade de que uma criança recém-nascida viva pelo menos 70 anos? Várias tentativas foram feitas, nenhuma delas totalmente aceitável, para dar interpretações alternativas a tais asserções probabilísticas.

Para a teoria matemática da probabilidade, a interpretação de probabilidades é irrelevante, exatamente como é irrelevante, na geometria, a interpretação de pontos, retas e planos. Usaremos a interpretação de freqüência relativa para probabilidades, apenas, como uma motivação intuitiva para as definições e teoremas que desenvolveremos ao longo do livro.

1.1. EXEMPLOS DE FENÔMENOS ALEATÓRIOS

Nesta seção discutiremos dois exemplos simples de fenômenos aleatórios com o objetivo de motivar a estrutura formal da teoria.

Exemplo 1. Uma caixa contém 5 bolas idênticas, porém numeradas de 1 a 5. Considere o seguinte experimento. As bolas são bem misturadas dentro da caixa e uma pessoa retira uma bola. Anota-se o número da bola, recolocando-a na caixa. O resultado do experimento é o número da bola selecionada. Não podemos fazer nenhuma previsão não-trivial sobre este experimento.

Suponha que repetimos n vezes o experimento acima. Denote por $N_n(k)$ o número de vezes que a bola de número k foi retirada nos n ensaios do experimento. Admita que tenhamos, $s = 3$ bolas e $n = 20$ ensaios. Os resultados destes 20 ensaios poderiam ser descritos listando os números que apareceram na ordem em que foram observados. Um resultado típico poderia ser

$$1, 1, 3, 2, 1, 2, 2, 3, 2, 3, 3, 2, 1, 2, 3, 3, 1, 3, 2, 2,$$

e neste caso teríamos

$$N_{20}(1) = 5, \qquad N_{20}(2) = 8, \qquad \text{e} \qquad N_{20}(3) = 7.$$

As freqüências relativas (isto é, proporção de vezes) dos resultados 1, 2 e 3 são então

$$\frac{N_{20}(1)}{20} = 0{,}25, \qquad \frac{N_{20}(2)}{20} = 0{,}40, \qquad \text{e} \qquad \frac{N_{20}(3)}{20} = 0{,}35.$$

À medida que o número de ensaios aumenta, espera-se que as freqüências relativas $N_n(1)/n, \ldots, N_n(s)/n$ se ajustem a alguns números fixos $p_1, p_2, \ldots, p_s$ (que, segundo nossa intuição, neste caso, deveriam ser iguais a $1/s$).

Pela interpretação de freqüência relativa, o número p_i seria a probabilidade de que a i-ésima bola seja retirada quando o experimento é realizado uma vez $(i = 1, 2, \ldots, s)$.

Construiremos agora um modelo matemático para o experimento de retirar uma bola da caixa. Para isto, tomamos primeiro um conjunto Ω contendo s pontos que colocamos em correspondência biunívoca com os possíveis resultados do experimento. Nesta correspondência exatamente um ponto de Ω estará associado com o resultado de que a bola com o número k seja selecionada. Chamemos este ponto ω_k. Associemos ao ponto ω_k o número $p_k = 1/s$ chamando-o de probabilidade de ω_k. Observamos de imediato que $0 \leqslant p_k \leqslant 1$ e que $p_1 + \cdots + p_s = 1$.

Suponha agora que além de serem numeradas de 1 a s, as r primeiras bolas, são pintadas de vermelho e, as $s - r$ restantes, são pintadas de preto. Realizamos o experimento como antes, mas agora estamos interessados apenas na cor da bola e não no seu número. Uma rápida reflexão mostra que a freqüência relativa de bolas vermelhas retiradas nas n repetições do experimento é simplesmente a soma das freqüências relativas $N_n(k)/n$, sobre os valores de k, que correspondem a bolas vermelhas. Esperaríamos, e a experiência confirma, que para n grande, esta freqüência relativa se ajustasse a algum número fixo. Como, para n grande, espera-se que as freqüências relativas $N_n(k)/n$ estejam próximas de $p_k = 1/s$, anteciparíamos que a freqüência relativa de bolas vermelhas se aproximaria de r/s. Novamente a experiência confirma este fato. Segundo a interpretação de freqüência relativa, chamaríamos então r/s de probabilidade de obter uma bola vermelha.

Vejamos como podemos incorporar este fato no nosso modelo. Seja A o subconjunto de Ω, consistindo daqueles pontos ω_k tais que a bola k é vermelha. Então A contém exatamente r pontos. Chamamos A um evento. De uma forma mais geral, nesta situação, chamaremos qualquer subconjunto B de Ω um evento. Dizer que o evento B ocorre, significa que o resultado do experimento é representado por algum ponto de B.

Sejam A e B dois eventos. Lembre-se que a união de A e B, $A \cup B$, é o conjunto de todos os pontos $\omega \in \Omega$ tais que $\omega \in A$ ou $\omega \in B$. Agora os pontos em Ω estão em correspondência com os resultados do nosso experimento. O evento A ocorre se o experimento produz um resultado que é representado por algum ponto em A, e analogamente o evento B ocorre se o resultado do experimento é representado por algum ponto em B. O conjunto $A \cup B$ representa, então, o fato que o evento A ocorre ou o evento B ocorre. De forma similar, a interseção $A \cap B$ de A e B consiste de todos os pontos que estão tanto em A quanto em B. Assim se $\omega \in A \cap B$ então $\omega \in A$ e $\omega \in B$ de modo que $A \cap B$ representa o fato de que ambos os eventos A e B ocorrem. O complemento A^c(ou A') de A é o conjunto de pontos em Ω que não estão em A. O evento A não ocorre se o experimento produz um resultado representado por um ponto em A^c.

Em um diagrama, se A e B são representados pelas regiões indicadas na Figura 1a, então $A \cup B$, $A \cap B$, e A^c são representados pelas regiões sombreadas nas Figuras 1b, 1c e 1d, respectivamente.

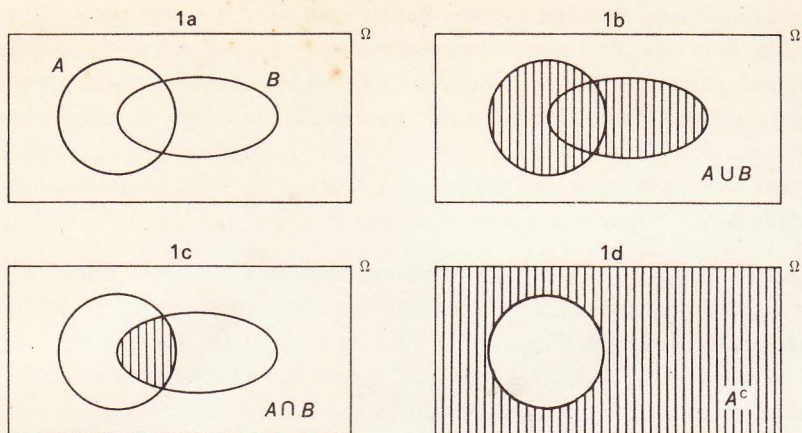



Figura 1

Para ilustrar estes conceitos seja A o evento "bola vermelha selecionada" e seja B o evento "bola selecionada com um número par". Então a união $A \cup B$ é o evento em que foi selecionada uma bola vermelha ou uma bola com número par. A interseção $A \cap B$ é o evento "seleção de bola vermelha com número par". O evento A^c ocorre se não foi selecionada uma bola vermelha.

Gostaríamos agora de associar probabilidades aos eventos. Matematicamente, isto significa simplesmente que associamos a cada conjunto B um número real. A priori poderíamos fazer isto de uma forma arbitrária. Entretanto, estaremos restringidos, se desejarmos que estas probabilidades reflitam o experimento que estamos tentando modelar. Como deveríamos fazer esta associação? Já associamos a cada ponto o número s^{-1}. Assim, a um conjunto de um único ponto $\{\omega\}$ deveria ser associado o número s^{-1}. Agora, de nossa discussão sobre a freqüência relativa do evento "extrair bola vermelha", parece que devemos associar ao evento A a probabilidade $P(A) = r/s$. De uma forma mais geral, se B é um evento qualquer, definiremos $P(B)$ através de $P(B) = j/s$, se B tem exatamente j pontos. Observamos então que

$$P(B) = \sum_{\omega_k \in B} p_k ,$$

onde $\Sigma_{\omega_k \in B}\, p_k$ significa que somamos os números p_k sobre os valores de k tais que $\omega_k \in B$. Da nossa definição de $P(B)$ segue-se facilmente que as afirmações seguintes são verdadeiras. Deixamos sua verificação para o leitor.

Seja ϕ o conjunto vazio; então $P(\phi) = 0$ e $P(\Omega) = 1$. Se A e B são dois conjuntos disjuntos quaisquer, isto é, $A \cap B = \phi$ então

$$P(A \cup B) = P(A) + P(B).$$

Exemplo 2. Sabe-se de experimentos físicos que um isótopo de uma certa substância é instável. Com o passar do tempo ele se degrada para uma forma mais estável através da emissão de nêutrons. Estamos interessados no tempo que um átomo de um isótopo leva para se degradar à forma estável. De acordo com as leis da física é impossível dizer com certeza quando um átomo específico se desintegrará, mas se observamos um número N de átomos, podemos fazer algumas previsões precisas sobre o número $N(t)$ de átomos que não se desintegram até o tempo t. Em outras palavras, podemos prever, com bastante precisão, a fração $N(t)/N$ de átomos que não se desintegram até o tempo t, mas não podemos dizer quais os átomos que permanecerão inalterados. Já que todos os átomos são idênticos, observar simultaneamente se N átomos seria equivalente a N repetições do mesmo experimento onde, neste caso, o experimento consiste em observar o tempo que um átomo leva para se desintegrar.

Em primeira aproximação (que é na realidade bastante precisa), a taxa com que o isótopo se desintegra, no tempo t, é proporcional ao número de átomos presentes no tempo t, de modo que $N(t)$ é dado aproximadamente pela solução da equação diferencial

$$\frac{df}{dt} = -\lambda f(t), \qquad f(0) = N,$$

onde $\lambda > 0$ é uma constante de proporcionalidade. A solução única desta equação é $f(t) = Ne^{-\lambda t}$, de modo que, a fração de átomos que não se desintegram até o tempo t, é dada aproximadamente por $N(t)/N = e^{-\lambda t}$. Se $0 \leqslant t_0 \leqslant t_1$, a fração de átomos que se desintegram no intervalo de tempo $[t_0, t_1]$ é $(e^{-\lambda t_0} - e^{-\lambda t_1})$. Conseqüentemente, de acordo com a interpretação de probabilidade como freqüência relativa, tomamos $(e^{-\lambda t_0} - e^{-\lambda t_1})$ como a probabilidade de que um átomo se desintegre entre os tempos t_0 e t_1.

Para fazer um modelo matemático deste experimento podemos tentar proceder como no exemplo anterior. Primeiro escolhemos um conjunto Ω que possa ser posto em correspondência um a um com os possíveis resultados do experimento. Um resultado, neste caso, é o tempo que um átomo leva para se desintegrar. Ele pode ser qualquer número real positivo, assim tomamos Ω como sendo o intervalo $[0, \infty)$ sobre o eixo dos números reais. De nossa discussão acima parece razoável associar a probabilidade $(e^{-\lambda t_0} - e^{-\lambda t_1})$ ao intervalo $[t_0, t_1]$. Em particular, se $t_0 = t_1 = t$, o intervalo se degenera no conjunto $\{t\}$ e a probabilidade associada a este conjunto é 0.

No exemplo anterior Ω tinha apenas um número finito de pontos; entretanto, aqui Ω tem um número infinito (não enumerável) de pontos e cada ponto tem probabilidade 0. Observamos novamente que $P(\Omega) = 1$ e $P(\phi) = 0$. Suponha que A e B sejam dois intervalos disjuntos. Então, a proporção de átomos que se desintegram no intervalo de tempo $A \cup B$, é a soma das proporções de átomos que se desintegram no intervalo A e no intervalo B. À luz desta aditividade exigimos

que no nosso modelo matemático $A \cup B$ tenha a probabilidade $P(A) + P(B)$ a ele associada. Em outras palavras, no nosso modelo matemático desejamos que

$$P(A \cup B) = P(A) + P(B)$$

sempre que A e B forem intervalos disjuntos.

1.2. ESPAÇOS DE PROBABILIDADE

Nosso propósito nesta seção é desenvolver uma estrutura matemática formal, chamada espaço de probabilidade, que forma a base para o tratamento matemático de fenômenos aleatórios.

Considere um experimento real ou imaginário que estamos procurando modelar. A primeira coisa que devemos fazer é decidir sobre os possíveis resultados do experimento. Não é muito sério se admitirmos, em nossa consideração, mais coisas do que realmente podem acontecer, mas desejamos estar certos de que não excluímos coisas que podem ocorrer. Uma vez decididos sobre os possíveis resultados, escolhemos um conjunto Ω cujos pontos ω estão associados a esses resultados. Entretanto, do ponto de vista estritamente matemático, Ω é simplesmente um conjunto abstrato de pontos.

Tomamos a seguir uma coleção não vazia $\mathscr{A}$ de subconjuntos de Ω, que representará a coleção de "eventos" aos quais desejamos associar probabilidades. Agora, um *evento* significa por definição um conjunto A em $\mathscr{A}$. A afirmação: *O evento A ocorre,* significa que o resultado do nosso experimento é representado por algum ponto $\omega \in A$. Novamente, do ponto de vista estritamente matemático, $\mathscr{A}$ é apenas uma coleção especificada de subconjuntos do conjunto Ω. Serão associadas probabilidades apenas aos conjuntos. $A \in \mathscr{A}$, isto é, eventos. No modelo do Exemplo 1, $\mathscr{A}$ consistia de todos os subconjuntos de Ω. Na situação geral em que Ω não tem um número finito de pontos, como no Exemplo 2, pode não ser possível escolher $\mathscr{A}$ desta maneira.

O problema seguinte é, o que deve ser a coleção $\mathscr{A}$? É bastante razoável exigir que $\mathscr{A}$ seja fechado sob uniões finitas e interseções finitas dos conjuntos em $\mathscr{A}$; bem como sob complementação. Por exemplo, se A e B são dois eventos, $A \cup B$ ocorre se o resultado do experimento é representado por um ponto em A ou em B. Claramente então, faz sentido falar sobre as probabilidades de que A e B ocorram, também deve fazer sentido falar sobre a probabilidade de que ou A ou B ocorra, isto é, de que o evento $A \cup B$ ocorra. Já que associaremos probabilidades somente aos conjuntos em $\mathscr{A}$, devemos exigir que $A \cup B \in \mathscr{A}$ sempre que A e B são membros de $\mathscr{A}$. Por outro lado $A \cap B$ ocorre se o resultado do experimento é representado por algum ponto que está em ambos os conjuntos A e B. Um raciocínio, análogo ao seguido para $A \cup B$, convence-nos que devemos ter $A \cap B \in \mathscr{A}$, sempre que $A, B \in \mathscr{A}$. Finalmente dizer que o evento A não ocorre é dizer que o resultado do experimento não é representado por um ponto em A, de modo que ele deve ser representado por algum ponto em A^c. Seria o

cúmulo da tolice dizer que podemos falar em probabilidade de A mas não na probabilidade de A^c. Assim exigiremos que A^c esteja em A sempre que $\mathscr{A}$ estiver em A^c.

Chegamos assim à conclusão de que $\mathscr{A}$ deve ser uma coleção não vazia de subconjuntos de Ω tendo as seguintes propriedades:

(i) Se A está em $\mathscr{A}$ também está em A^c.

(ii) Se A e B estão em $\mathscr{A}$, $A \cup B$ e $A \cap B$ também estão

Um simples argumento indutivo mostra que se $A_1, A_2, \ldots, A_n$ são conjuntos em $\mathscr{A}$, $\bigcup_{i=1}^n A_i$ e $\bigcap_{i=1}^n A_i$ também o são. Aqui, usamos a notação abreviada

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n$$

e

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n.$$

Já que $A \cap A^c = \phi$ e $A \cup A^c = \Omega$, vemos também que o conjunto vazio e o conjunto Ω devem estar em $\mathscr{A}$.

Uma coleção não-vazia de subconjuntos, de um dado conjunto Ω, que é fechada sob finitas operações da teoria dos conjuntos, é chamada *álgebra de subconjuntos* de Ω. Portanto, parece que devemos exigir que $\mathscr{A}$ seja uma álgebra de subconjuntos. Acontece, entretanto, que devido a certas razões matemáticas, é insuficiente tomar $\mathscr{A}$ como sendo uma álgebra de subconjuntos. O que realmente exigiremos da coleção $\mathscr{A}$ é mais restritivo. Exigiremos que $\mathscr{A}$ seja fechada não somente sob finitas operações da teoria dos conjuntos, mas também, sob um número infinito enumerável de operações da teoria dos conjuntos. Em outras palavras, se $\{A_n\}$, $n \geqslant 1$, é uma seqüência de conjuntos em $\mathscr{A}$, exigiremos que

$$\bigcup_{n=1}^{\infty} A_n \in \mathscr{A} \quad \text{e} \quad \bigcap_{n=1}^{\infty} A_n \in \mathscr{A}.$$

Usamos aqui a notação

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup A_2 \cup \cdots$$

para representar a união de todos os conjuntos da seqüência e

$$\bigcap_{n=1}^{\infty} A_n = A_1 \cap A_2 \cap \cdots$$

para representar a interseção de todos os conjuntos de seqüências. Uma coleção de subconjuntos, de um dado conjunto Ω, que é fechada sob um número infinito enumerável de operações da teoria dos conjuntos é chamada σ-álgebra de subconjuntos de Ω. (Usa-se σ para distinguir uma tal coleção de uma álgebra de subconjuntos.) De um modo mais formal, temos a seguinte:

Definição 1. Diz-se que uma coleção não-vazia $\mathscr{A}$ de subconjunto de Ω é uma σ-álgebra de subconjuntos de Ω, desde que, as seguintes propriedades sejam satisfeitas:

(i) Se A está em $\mathscr{A}$, A^c também está em $\mathscr{A}$.
(ii) Se A_n está em $\mathscr{A}$, $n = 1, 2, \ldots$, então $\bigcup_{n=1}^{\infty} A_n$ e $\bigcap_{n=1}^{\infty} A_n$ também estão em $\mathscr{A}$.

Chegamos agora ao problema de associar probabilidades aos eventos. Deixamos claro nos exemplos da seção anterior que a probabilidade de um evento é um número real não-negativo. Para um evento A, seja $P(A)$ a sua probabilidade. Então $0 \leqslant P(A) \leqslant 1$. Ao conjunto Ω representando todos os resultados possíveis deve, naturalmente, ser associado o número 1, de modo que $P(\Omega) = 1$. Mostramos na discussão do Exemplo 1 que se A e B são dois eventos disjuntos quaisquer, $P(A \cup B) = P(A) + P(B)$. De forma semelhante, mostramos no Exemplo 2 que se A e B são dois intervalos disjuntos, devíamos exigir também que

$$P(A \cup B) = P(A) + P(B).$$

Parece então razoável exigir, em geral, que se A e B são dois eventos disjuntos, então $P(A \cup B) = P(A) + P(B)$. Se seguiria então por indução que se $A_1, A_2, \ldots, A_n$ são n conjuntos mutuamente disjuntos (isto é, se $A_i \cap A_j = \phi$ sempre que $i \neq j$), então

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

Na realidade, novamente por razões matemáticas, exigimos que esta propriedade aditiva se verifique para coleções enumeráveis de eventos disjuntos.

Definição 2. Uma medida de probabilidade P, sobre um σ-álgebra de um subconjunto $\mathscr{A}$ de um conjunto Ω, é uma função real cujo domínio é $\mathscr{A}$ e que satisfaz as seguintes probabilidades:

(i) $P(\Omega) = 1$.
(ii) $P(A) \geqslant 0$ para todo $A \in \mathscr{A}$.
(iii) Se A_n, $n = 1, 2, 3, \ldots$, são conjuntos mutuamente disjuntos em $\mathscr{A}$, então

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Um espaço de probabilidade, representado por $(\Omega, \mathscr{A}, P)$, é um conjunto Ω, um σ-álgebra de subconjuntos $\mathscr{A}$, e uma medida de probabilidade P definida em $\mathscr{A}$.

É bastante fácil encontrar um espaço de probabilidade que corresponda ao experimento de extrair uma bola de uma caixa, e este, já foi dado, em essência,

na discussão deste experimento. Simplesmente tomamos Ω como sendo um conjunto finito contendo s pontos, $\mathscr{A}$ como sendo a coleção de todos os subconjuntos de Ω e P como sendo a medida de probabilidade que associa a probabilidade $P(A) = j/s$ ao evento A se A contém exatamente j pontos.

Consideremos agora o espaço de probabilidade associado ao experimento da desintegração de isótopo (Exemplo 2). Neste caso, é claro que $\Omega = [0, \infty)$, mas não é tão óbvio o que $\mathscr{A}$ e P devem ser. Na verdade, como indicaremos abaixo, este não é, de modo algum, um problema trivial, e sim um problema em que todas as suas ramificações dependem de algumas propriedades da teoria dos conjuntos que estão além do escopo deste livro.

Uma coisa no entanto é clara: quaisquer que sejam as escolhas de $\mathscr{A}$ e P, $\mathscr{A}$ deve conter todos os intervalos e P deve associar a probabilidade $(e^{-\lambda t_0} - e^{-\lambda t_1})$ ao intervalo $[t_0, t_1]$, se desejamos que o espaço de probabilidade que estamos construindo, reflita a situação física. Então o problema de construir um espaço transforma-se no seguinte problema, puramente matemático: Existe um σ-álgebra $\mathscr{A}$ que contenha todos os intervalos e uma medida de probabilidade P definida em $\mathscr{A}$ que associa a probabilidade desejada $P(A)$ ao intervalo A? Problemas deste tipo estão no domínio de um ramo da matemática avançada chamada teoria da medida e não podem ser tratados ao nível deste livro. Resultados da teoria da medida mostram que a resposta a este problema particular e a outros de mesma natureza é afirmativa, tais construções são sempre possíveis.

Não nos deteremos em construções de espaços de probabilidades em geral. A teoria matemática da probabilidade começa com um espaço abstrato de probabilidade e desenvolve a teoria usando o espaço de probabilidade como uma base de operação. Executando a formação de uma base para definir precisamente outros conceitos envolvidos na teoria, o espaço de probabilidade desempenha um papel muito pequeno no desenvolvimento subseqüente da teoria. Quantidades auxiliares (especialmente variáveis aleatórias, um conceito abordado no Capítulo 3) se transformam rapidamente em tema dominante da teoria e, o espaço de probabilidade em si, recua para um plano secundário.

Concluiremos a discussão de espaços de probabilidade construindo uma importante classe de espaços de probabilidade denominados *espaços uniformes de probabilidade*.

Alguns problemas mais antigos em probabilidade envolvem a idéia de escolher "ao acaso" um ponto de um conjunto S. Nossas idéias intuitivas sobre esta noção nos mostram que se A e B são conjuntos de mesmo "tamanho", a chance de escolher um ponto de A deve ser a mesma que a de escolher de B. Se S tem apenas um número finito de pontos, podemos medir o tamanho de um conjunto por meio de sua cardinalidade. Assim dois conjuntos são do mesmo "tamanho" se contêm o mesmo número de pontos. É bastante fácil construir um espaço de probabilidade correspondente ao experimento de escolher ao acaso um ponto de um conjunto S. Tomamos $\Omega = S$ e $\mathscr{A}$ como sendo todos os subconjuntos de S,

e associamos ao conjunto A a probabilidade $P(A) = j/s$ se A é um conjunto que contém exatamente j pontos. Tal espaço de probabilidade é chamado *espaço simétrico de probabilidade* porque cada conjunto com um ponto tem a mesma probabilidade s^{-1}. Voltaremos ao estudo de tais espaços no Capítulo 2.

Suponha agora que S é o intervalo $[a, b]$ sobre o eixo real, onde $-\infty < a < b < +\infty$. Neste caso, parece razoável medir o "tamanho" de um subconjunto A de $[a, b]$ através do seu comprimento. Então, dois conjuntos são do mesmo tamanho se tiverem o mesmo comprimento. Representaremos o comprimento de um conjunto A por $|A|$.

Para construir um espaço de probabilidade para o experimento de "escolher ao acaso um ponto S" procedemos de maneira semelhante àquela adotada para o experimento do isótopo. Tomamos $\Omega = S$ e lançamos mão dos resultados da teoria da medida, que mostram que existe um σ-álgebra $\mathscr{A}$ de subconjuntos de S e uma medida de probabilidade P definida em $\mathscr{A}$ tal que $P(A) = |A|/|S|$ sempre que A é um intervalo.

De uma forma mais geral, seja S um subconjunto qualquer do espaço Euclidiano r-dimensional, tendo um volume r-dimensional finito e não nulo. Seja $|A|$ o volume de um subconjunto A de S. Então, existe um σ-álgebra $\mathscr{A}$ de subconjuntos de S que contém todos os subconjuntos de S que possuem volumes a eles associados como em cálculo e uma medida de probabilidade P definida em $\mathscr{A}$, tal que $P(A) = |A|/|S|$ para qualquer conjunto A. Tal espaço será designado *espaço uniforme de probabilidade* e representado por $(S, \mathscr{A}, P)$.

1.3. PROPRIEDADES DAS PROBABILIDADES

Derivaremos nesta seção algumas propriedades adicionais de uma medida de probabilidade P que decorrem de sua própria definição. Estas propriedades serão usadas constantemente ao longo do restante deste livro. Assumimos que seja dado algum espaço de probabilidades $(\Omega, \mathscr{A}, P)$ e que todos os conjuntos em discussão são eventos, isto é, membros de $\mathscr{A}$.

Para um conjunto qualquer A temos $A \cup A^c = \Omega$ e assim para dois conjuntos quaisquer A e B temos a decomposição de B:

(1) $$B = \Omega \cap B = (A \cup A^c) \cap B = (A \cap B) \cup (A^c \cap B).$$

Uma vez que $A \cap B$ e $A^c \cap B$ são disjuntos, vemos que de (iii) da Definição 2 que

(2) $$P(B) = P(A \cap B) + P(A^c \cap B).$$

Fazendo $B = \Omega$ e lembrando que $P(\Omega) = 1$, concluímos de (2) que

(3) $$P(A^c) = 1 - P(A).$$

Em particular $P(\phi) = 1 - P(\Omega)$, de modo que

(4) $$P(\phi) = 0.$$

Como uma segunda aplicação de (2), suponha que $A \subset B$. Então $A \cap B = A$ e portanto

(5) $$P(B) = P(A) + P(A^c \cap B) \text{ se } A \subset B.$$

Já que $P(A^c \cap B) \geqslant 0$ em virtude de (ii), vemos de (5) que

(6) $$P(B) \geqslant P(A) \text{ se } A \subset B.$$

As leis de De Morgan estabelecem que se $\{A_n\}$, $n \geqslant 1$, é uma seqüência qualquer de conjuntos, então

(7) $$\left(\bigcup_n A_n\right)^c = \left(\bigcap_n A_n^c\right)$$

e

(8) $$\left(\bigcap_n A_n\right)^c = \left(\bigcup_n A_n^c\right).$$

Para ver que (7) é verdadeiro, observe que $\omega \in (\bigcup_{n \geq 1} A_n)^c$ se, e somente se, $\omega \notin A_n$ para qualquer n, isto é, $\omega \in A_n^c$ para todo $n \geqslant 1$, ou equivalentemente $\omega \in \bigcap_n A_n^c$. Para estabelecer (8) aplicamos (7) a $\{A_n^c\}$, obtendo

$$\left(\bigcup_n A_n^c\right)^c = \bigcap_n A_n,$$

e tomando o complemento vemos que

$$\bigcup_n A_n^c = \left(\bigcap_n A_n\right)^c.$$

Uma relação útil que decorre de (7) e (3) é

(9) $$P\left(\bigcup_n A_n\right) = 1 - P\left(\bigcap_n A_n^c\right).$$

Mas $\bigcup_n A_n$ é o evento de que pelo menos um dos eventos A_n ocorre, enquanto $\bigcap_n A_n^c$ é o evento de que nenhum desses eventos ocorre. Em palavras, (9) afirma que a probabilidade de que pelo menos um dos eventos A_n ocorra é 1 menos a probabilidade de que nenhum dos eventos A_n ocorra. A vantagem de (9) é que em algumas situações é mais fácil determinar $P(\bigcap_n A_n^c)$ do que $P(\bigcup_n A_n)$. [Note que desde que os eventos A_n não são necessariamente disjuntos, não é verdadeiro que $P(\bigcup_n A_n) = \sum_n P(A_n)$.] O exemplo a seguir ilustra com propriedade o uso da expressão (9).

Exemplo 3. Suponha que se lance três moedas idênticas e perfeitamente equilibradas. Determine a probabilidade de obter pelo menos uma cara.

Representando cara por H e coroa por T, existem oito resultados possíveis para este experimento

MOEDA 1	H	H	H	H	T	T	T	T
MOEDA 2	H	H	T	T	H	H	T	T
MOEDA 3	H	T	H	T	H	T	H	T

A intuição sugere que cada um desses resultados deve ter a probabilidade de ocorrência 1/8. Seja A_1 o evento de que a primeira moeda apresenta cara, A_2 o evento de que a segunda moeda apresenta cara e A_3 o evento de que a terceira moeda apresenta cara. O problema pede a determinação de $P(A_1 \cup A_2 \cup A_3)$. Porém $A_1^c \cap A_2^c \cap A_3^c = \{ T, T, T \}$ e assim

$$P(A_1^c \cap A_2^c \cap A_3^c) = 1/8;$$

portanto (9) implica que

$$P(A_1 \cup A_2 \cup A_3) = 1 - P(A_1^c \cap A_2^c \cap A_3^c) = 7/8.$$

O postulado básico (iii) sobre medidas de probabilidade diz-nos que $P(A \cup B) = P(A) + P(B)$, para conjuntos disjuntos A e B. Se A e B não são necessariamente disjuntos, então

(10) $$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

e conseqüentemente

(11) $$P(A \cup B) \leqslant P(A) + P(B).$$

Para ver que (10) é verdadeiro, observe que os conjuntos $A \cap B^c$, $A \cap B$, $A^c \cap B$ são mutuamente disjuntos e sua união é simplesmente $A \cup B$ (ver Figura 2). Assim

(12) $$P(A \cup B) = P(A \cap B^c) + P(A^c \cap B) + P(A \cap B).$$

Entretanto, em virtude de (2)

$$P(A \cap B^c) = P(A) - P(A \cap B)$$

e

$$P(A^c \cap B) = P(B) - P(A \cap B).$$

Substituindo essas expressões em (12), obtemos (10).

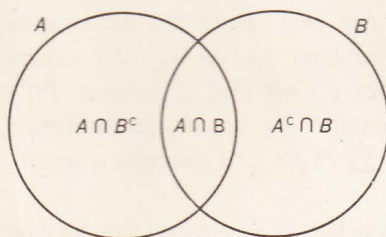



Figura 2

As expressões (10) e (11) estendem-se para qualquer número finito de conjuntos. O análogo da fórmula exata (10) é um tanto complicado e será discutido no Capítulo 2. Entretanto, a desigualdade (11) pode ser estendida facilmente por indução, obtendo-se

(13) $$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{i=1}^{n} P(A_i).$$

Para demonstrá-lo, observe que se $n \geqslant 2$, então por (11)

$$P(A_1 \cup \cdots \cup A_n) = P((A_1 \cup \cdots \cup A_{n-1}) \cup A_n)$$
$$\leq P(A_1 \cup \cdots \cup A_{n-1}) + P(A_n).$$

Portanto se (13) é verdadeiro para $n - 1$ conjuntos, também o é para n conjuntos. Já que (13) é claramente verdadeiro para $n = 1$, o resultado fica demonstrado por indução.

Até aqui usamos o fato de que uma medida de probabilidade é finitamente aditiva. O resultado seguinte usa a aditividade enumerável.

Teorema 1. Sejam os eventos $A_n, n \geqslant 1$.

(i) Se $A_1 \subset A_2 \subset \cdots$ e $A = \bigcup_{n=1}^{\infty} A_n$, então

(14) $$\lim_{n \to \infty} P(A_n) = P(A).$$

(ii) Se $A_1 \supset A_2 \supset \cdots$ e $A = \bigcap_{n=1}^{\infty} A_n$, então (14) se verifica também.

Demonstração de (i). Suponha que $A_1 \subset A_2 \subset \cdots$ e $A = \bigcup_{n=1}^{\infty} A_n$. Seja $B_1 = A_2$ e para todo $n \geqslant 2$, seja B_n o conjunto de pontos que estão em A_n mas não em A_{n-1}, isto é, $B_n = A_n \cap A_{n-1}^c$. Um ponto ω está em B_n se e somente se, ω está em A e A_n é o primeiro conjunto da seqüência $A_1, A_2, \ldots$ que contém ω. Por definição os conjuntos B_n são disjuntos,

$$A_n = \bigcup_{i=1}^{n} B_i,$$

e

$$A = \bigcup_{i=1}^{\infty} B_i.$$

Conseqüentemente

$$P(A_n) = \sum_{i=1}^{n} P(B_i)$$

e

$$P(A) = \sum_{i=1}^{\infty} P(B_i).$$

Mas

(15) $$\lim_{n \to \infty} \sum_{i=1}^{n} P(B_i) = \sum_{i=1}^{\infty} P(B_i)$$

de acordo com a definição da soma de uma série infinita. De (15), segue-se que

$$\lim_{n \to \infty} P(A_n) = \lim_{n \to \infty} \sum_{i=1}^{n} P(B_i)$$
$$= \sum_{i=1}^{\infty} P(B_i) = P(A),$$

de modo que (14) é verdadeiro.

Demonstração de (ii). Suponha que $A_1 \supset A_2 \supset \cdots$ e $A = \bigcap_{n=1}^{\infty} A_n$. Então $A_1^c \subset A_2^c \subset \cdots$ e em virtude de (8)

$$A^c = \bigcup_{n=1}^{\infty} A_n^c.$$

Assim pelo item (i) do teorema

$$\lim_{n \to \infty} P(A_n^c) = P(A^c) \tag{16}$$

Como $P(A_n^c) = 1 - P(A_n)$ e $P(A^c) = 1 - P(A)$, segue-se de (16) que

$$\begin{aligned} \lim_{n \to \infty} P(A_n) &= \lim_{n \to \infty} (1 - P(A_n^c)) \\ &= 1 - \lim_{n \to \infty} P(A_n^c) \\ &= 1 - P(A^c) = P(A), \end{aligned}$$

e novamente (14) se verifica.

1.4. PROBABILIDADE CONDICIONAL

Considere uma caixa contendo r bolas vermelhas numeradas de 1 a r e b bolas pretas numeradas de 1 a b. Suponha que a probabilidade de extrair qualquer bola é $(b + r)^{-1}$. Se sabemos que a bola extraída é vermelha, qual a probabilidade de que seu número seja 1? Outra maneira de formular este problema é como segue. Seja A o evento de que a bola selecionada é vermelha e seja B o evento de que o número da bola selecionada é um. O problema então é determinar a probabilidade do evento B ter ocorrido, dado que ocorreu o evento A. Este problema não pode ser resolvido sem ter uma definição precisa da probabilidade condicional de um evento, dado um outro evento. Esta definição é a seguinte:

Definição 3. Sejam dois eventos A e B tais que $P(A) > 0$. Então define-se a probabilidade condicional de B dado A, representada por $P(B \mid A)$, como sendo

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)}. \tag{17}$$

Se $P(A) = 0$, a probabilidade de B dado A é indefinido.

A definição acima pode ser facilmente motivada pela interpretação de probabilidades como freqüências relativas. Considere um experimento que é repetitivo um grande número de vezes. Sejam $N_n(A)$, $N_n(B)$ e $N_n(A \cap B)$ o número de vezes que os eventos A, B e $A \cap B$ ocorrem em n repetições do experimento. Se registrássemos somente os experimentos em que A ocorre, teríamos $N_n(A)$

provas nas quais B ocorre $N_n(A \cap B)$ vezes. Assim, a proporção de vezes que B ocorre nestes $N_n(A)$ experimentos é $N_n(A \cap B)/N_n(A)$. Mas

$$\frac{N_n(A \cap B)}{N_n(A)} = \frac{N_n(A \cap B)/n}{N_n(A)/n}$$

e assim para valores grandes de n esta fração deve estar próxima de $P(A \cap B)/P(A)$.

Como primeiro exemplo do uso de (17) resolvemos o problema proposto no começo desta seção. Já que Ω contém $b + r$ pontos, cada um com probabilidade $(b + r)^{-1}$ vemos que $P(A) = r(b + r)^{-1}$ e $P(A \cap B) = (b + r)^{-1}$. Assim

$$P(B \mid A) = \frac{1}{r}.$$

Esta probabilidade deve ser comparada com a probabilidade "incondicional" de B que é $P(B) = 2(b + r)^{-1}$.

Exemplo 4. Considere o lançamento de duas moedas idênticas e perfeitamente equilibradas.

(a) Determine a probabilidade condicional de obter duas caras, dado que se obteve cara na primeira moeda.

(b) Determine a probabilidade condicional de obter duas caras, dado que se obteve pelo menos uma cara.

Para resolver estes problemas, tomamos o espaço de amostra Ω consistindo de quatro pontos HH, HT, TH, TT, cada um com probabilidade 1/4. Seja A o evento de obter cara na primeira moeda e B o de obter cara na segunda. Para resolver (a) determinamos

$$P(A \cap B \mid A) = P(A \cap B)/P(A) = (1/4)(1/2) = 1/2.$$

Para resolver (b) determinamos

$$P(A \cap B \mid A \cup B) = P(A \cap B)/P(A \cup B) = (1/4)/(3/4) = 1/3.$$

Nos exemplos acima, o espaço de probabilidade era especificado e usamos (17) para determinar diversas probabilidades condicionais. Entretanto, em muitos problemas procedemos realmente na direção oposta. Partindo do conhecimento antecipado de valores que algumas probabilidades condicionais devem assumir, usamos essa informação para determinar a medida de probabilidade em Ω. Apresentamos a seguir um exemplo típico dessa situação.

Exemplo 5. Suponha que a população de uma certa cidade é constituída por 40% de homens e 60% de mulheres. Suponha ainda que 50% dos homens e 30% das mulheres são fumantes. Determine a probabilidade de que uma pessoa que fuma seja homem.

Representamos por M o evento de que a pessoa selecionada é homem e por F o evento de que a pessoa selecionada é mulher. Seja S o evento de que a

pessoa selecionada é fumante e por N o de que a pessoa não é fumante. Então os dados do problema são: $P(S \mid M) = 0{,}5; P(S \mid F) = 0{,}3; P(M) = 0{,}4$ e $P(F) = 0{,}6$. O problema consiste em determinar $P(M \mid S)$. De acordo com (17).

$$P(M \mid S) = \frac{P(M \cap S)}{P(S)}.$$

Mas $P(M \cap S) = P(M)\, P(S \mid M) = (0{,}4)(0{,}5) = 0{,}20$, de modo que o numerador pode ser determinado em termos de probabilidades conhecidas. Como S é a união de dois conjuntos disjuntos $S \cap M$ e $S \cap F$, segue-se que

$$P(S) = P(S \cap M) + P(S \cap F)$$

Já que

$$P(S \cap F) = P(F)\, P(S \mid F) = (0{,}6)(0{,}3) = 0{,}18,$$

Vemos que

$$P(S) = 0{,}20 + 0{,}18 = 0{,}38$$

Assim

$$P(M \mid S) = \frac{0{,}20}{0{,}38} \approx 0{,}53.$$

O leitor observará que o espaço de probabilidade nunca foi mencionado explicitamente como tal. Resolve-se este problema e outros de natureza similar usando a informação dada e as regras para determinar probabilidades dadas na Seção 3 para obter as probabilidades desejadas.

É bastante fácil construir um espaço de probabilidade para o exemplo acima. Toma-se para Ω o conjunto formado de quatro pontos SM, SF, NM e NF que são os únicos pontos nos conjuntos $S \cap M$, $S \cap F$, $N \cap M$ e $N \cap F$, respectivamente. As probabilidades associadas a esses pontos não são especificadas diretamente, mas precisam ser determinadas de tal forma que os eventos $(S \mid M), (S \mid F)$, M e F tenham as probabilidades especificadas. Já determinamos $P(S \cap M) = 0{,}20$ e $P(S \cap F) = 0{,}18$. Deixamos como exercício a determinação das probabilidades associadas aos outros dois pontos.

O problema discutido neste exemplo é um caso especial da situação geral que passamos a considerar. Suponha que $A_1, A_2, \ldots, A_n$ são n conjuntos mutuamente disjuntos cuja união é Ω. Seja B em evento tal que $P(B) > 0$ e suponha que $P(B \mid A_k)$ e $P(A_k)$ são conhecidas para $1 \leqslant k \leqslant n$. Qual é o valor de $P(A_i \mid B)$? Para resolver este problema, observe que A_k são conjuntos disjuntos cuja união é Ω, de modo que

$$B = B \cap \left(\bigcup_{k=1}^{n} A_k \right) = \bigcup_{k=1}^{n} (B \cap A_k).$$

Assim

$$P(B) = \sum_{k=1}^{n} P(B \cap A_k).$$

Mas

$$P(B \cap A_k) = P(A_k)\, P(B \mid A_k)$$

de modo que podemos escrever

$$(18) \qquad P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B \mid A_i)}{\sum_{k=1}^{n} P(A_k)P(B \mid A_k)}.$$

Esta fórmula, chamada *regra de Bayes*, tem aplicação freqüente. Uma forma de interpretar o resultado (18) é a seguinte: Suponha que pensemos nos eventos A_k como as possíveis "causas" do evento observável B. Então $P(A_i \mid B)$ é a probabilidade de que o evento A_i foi a causa de B, dado que B ocorreu. A regra de Bayes também forma a base de um método estatístico chamado método Bayesiano que será discutido no Volume II, *Introdução à Teoria Estatística.*

Como ilustração da regra de Bayes, consideramos o problema seguinte (já meio clássico).

Exemplo 6. Suponha que existam três cofres, cada um com duas gavetas. O primeiro tem uma moeda de ouro em cada gaveta, o segundo tem uma moeda de ouro em uma gaveta e uma moeda de prata em outra, e o terceiro cofre tem uma moeda de prata em cada gaveta. Escolhe-se um cofre ao acaso e abre-se uma gaveta. Se a gaveta contém uma moeda de ouro, qual a probabilidade de que a outra gaveta contenha também uma moeda de ouro? Pedimos ao leitor que faça uma pausa e adivinhe a resposta antes de ler a solução. Freqüentemente a resposta errada de 1/2 é dada para este problema.

Resolve-se o problema fácil e corretamente usando a regra de Bayes uma vez decifrada a descrição. Podemos pensar em um espaço de probabilidade em que os eventos A_1, A_2 e A_3 correspondem às seleções do primeiro, segundo e terceiro cofre, respectivamente. Estes eventos são disjuntos e sua união é Ω, já que se seleciona exatamente um cofre. Além do mais, presume-se que os três cofres são igualmente prováveis de serem selecionados, de modo que $P(A_i) = 1/3$, $i = 1, 2, 3$. Seja B o evento de que a moeda observada é de ouro. Então, da composição dos cofres é claro que

$$P(B \mid A_1) = 1, \qquad P(B \mid A_2) = 1/2 \qquad \text{e} \qquad P(B \mid A_3) = 0$$

O problema pede a probabilidade de que a segunda gaveta contenha uma moeda de ouro, dado que havia uma moeda de ouro na primeira. Isto pode acontecer somente se o cofre escolhido foi o primeiro, assim o problema equivale ao de determinar $P(A_1 \mid B)$. Agora podemos aplicar a regra de Bayes (18) para obter a resposta que é 2/3. Deixamos ao leitor como exercício a determinação de probabilidade de que a segunda gaveta contenha uma moeda de prata, dado que a primeira continha uma de ouro.

Para exemplo seguinte consideramos um esquema simples de probabilidade devido a Polya.

Exemplo 7. Esquema de urna de Polya. Suponha que uma urna contenha r bolas vermelhas e b bolas pretas. Extrai-se uma bola e observa-se a sua cor. A seguir coloca-se na urna a bola extraída juntamente com $c > 0$ bolas da mesma cor. Este procedimento é repetido $(n - 1)$ vezes mais, de modo que o número total de extrações é n.

Seja R_j, $1 \leqslant j \leqslant n$, o evento de que a j-ésima bola selecionada é vermelha e seja B_j, $1 \leqslant j \leqslant n$, o evento de que a j-ésima bola selecionada é preta. Naturalmente R_j e B_j são disjuntos para um dado j. No momento da k-ésima extração existem $b + r + (k - 1)c$ bolas na urna e assumimos que a probabilidade de selecionar qualquer bola particular é $(b + r + (k - 1)c)^{-1}$. Para determinar $P(R_1 \cap R_2)$ escrevemos

$$P(R_1 \cap R_2) = P(R_1)\, P(R_2 \mid R_1).$$

Mas

$$P(R_1) = \frac{r}{b + r}, \qquad P(R_2 \mid R_1) = \frac{r + c}{b + r + c},$$

e assim

$$P(R_1 \cap R_2) = \left(\frac{r}{b + r}\right)\left(\frac{r + c}{b + r + c}\right).$$

De forma similar

$$P(B_1 \cap R_2) = \left(\frac{b}{b + r}\right)\left(\frac{r}{b + r + c}\right)$$

e assim

$$\begin{aligned} P(R_2) &= P(R_1 \cap R_2) + P(B_1 \cap R_2) = \\ &= \left(\frac{r}{b + r}\right)\left(\frac{r + c}{b + r + c}\right) + \left(\frac{b}{b + r}\right)\left(\frac{r}{b + r + c}\right) = \\ &= \frac{r}{b + r}. \end{aligned}$$

Conseqüentemente $P(R_2) = P(R_1)$. Já que

$$P(B_2) = 1 - P(R_2) = \frac{b}{b + r},$$

temos $P(B_2) = P(B_1)$. Propriedades adicionais do esquema de Polya serão desenvolvidas nos exercícios.

1.5. INDEPENDÊNCIA

Considere uma caixa contendo quatro bolas distintas e um experimento que consiste em extrair uma bola da caixa. Assumimos que a extração de qualquer bola é igualmente provável. Seja $\Omega = \{1, 2, 3, 4\}$, cada ponto com probabilidade $1/4$.

Sejam dois eventos A e B. Para certas escolhas de A e B, o conhecimento de que A ocorre, aumenta a chance de B ocorrer. Por exemplo, se $A = \{1, 2\}$ e $B = \{1\}$, então $P(A) = 1/2$, $P(B) = 1/4$ e $P(A \cap B) = 1/4$. Conseqüentemente $P(B \mid A) = 1/2$, que é maior que $P(B)$. Por outro lado, para outras escolhas de A e B, o conhecimento de que A ocorre, diminui a chance de B ocorrer. Por exemplo: se $A = \{1, 2, 3\}$, $B = \{1, 2, 4\}$, então $P(A) = 3/4$, $P(B) = 3/4$ e $P(A \cap B) = 1/2$. Portanto $P(B \mid A) = 2/3$ que é menor que $P(B)$.

Um caso muito interessante ocorre quando o conhecimento de que A ocorre não altera a chance de ocorrência de B. Como um exemplo disso, seja $A = \{1, 2\}$ e $B = \{1, 3\}$; então $P(A) = 1/2$, $P(B) = 1/2$ e $P(A \cap B) = 1/4$, e portanto $P(B \mid A) = 1/2$. Eventos como esses, para os quais as probabilidades condicional e incondicional são iguais, são chamados eventos independentes.

Sejam A e B dois eventos quaisquer em um espaço geral de probabilidade, e suponha que $P(A) \neq 0$. Podemos definir os eventos A e B como sendo independentes se $P(B \mid A) = P(B)$. Como $P(B \mid A) = P(B \cap A)/P(A)$, vemos que se A e B são independentes, então

$$P(A \cap B) = P(A)P(B). \tag{19}$$

Como (19) faz sentido mesmo que $P(A) = 0$ e é também simétrica em A e B, ela conduz a uma definição alternativa de independência.

Definição 4. Dois eventos A e B são independentes se e somente se,

$$P(A \cap B) = P(A)P(B).$$

Podemos considerar um problema semelhante para três conjuntos A, B e C. Seja $\Omega = \{1, 2, 3, 4\}$, cada ponto com probabilidade $1/4$. Seja $A = \{1, 2\}$, $B = \{1, 3\}$ e $C = \{1, 4\}$. Deixamos como exercício, mostrar que os pares de eventos A e B, A e C e B e C são independentes. Dizemos que os eventos A, B e C são independentes *aos pares*. Por outro lado, $P(C) = 1/2$ e

$$P(C \mid A \cap B) = 1.$$

Assim, o conhecimento de que o evento $A \cap B$ ocorre, aumenta a chance de C ocorrer. Neste sentido os eventos A, B e C não chegam a ser *mutuamente* independentes. Em geral três eventos A, B e C são mutuamente independentes se são independentes aos pares e se

$$P(A \cap B \cap C) = P(A)P(B)P(C).$$

Como exercício, mostre que se A, B e C são mutuamente independentes e $P(A \cap B) \neq 0$, então $P(C \mid A \cap B) = P(C)$.

De um modo geral, definimos $n \geqslant 3$ eventos $A_1, A_2, \ldots, A_n$ como sendo mutuamente independentes se

$$P(A_1 \cap A_2 \cap \cdots A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

e se qualquer subcoleção contendo pelo menos dois porém, menos que n eventos são mutuamente independentes.

Exemplo 8. Seja S o quadrado no plano $0 \leqslant x \leqslant 1$ e $0 \leqslant y \leqslant 1$. Considere o espaço uniforme de probabilidade sobre o quadrado, e seja A o evento.

$$\{(x, y): 0 \leq x \leq 1/2, 0 \leq y \leq 1\}$$

e B o evento

$$\{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1/4\}.$$

Note que A e B são eventos independentes.

Para tanto, determinamos $P(A)$, $P(B)$ e $P(A \cap B)$ e mostramos que $P(A \cap B) = P(A)\,P(B)$. Como A é um sub-retângulo do quadrado S com área $1/2$ e B é um sub-retângulo de S com área $1/4$, segue-se que $P(A) = 1/2$ e $P(B) = 1/4$. Por outro lado

$$A \cap B = (x, y): 0 \leqslant x \leqslant 1/2, 0 \leqslant y \leqslant 1/4$$

é um sub-retângulo do quadrado S com área $1/8$. Assim $P(A \cap B) = 1/8$ e vemos que A e B são eventos independentes.

Usa-se freqüentemente a noção de independência para construir espaços de probabilidades correspondentes a repetições de um mesmo experimento. Daremos um tratamento mais completo a esse assunto no Capítulo 3. Aqui nos contentaremos em examinar a situação mais simples, aquela que envolve experimentos (como o lançamento de uma moeda possivelmente não tendenciosa) que podem conduzir a apenas um dos dois resultados possíveis – sucesso ou fracasso.

Em um experimento com n lançamentos de uma moeda, onde sucesso e fracasso em cada lançamento ocorrem com probabilidades p e $1 - p$, respectivamente, acreditamos intuitivamente que o resultado do i-ésimo lançamento não deve ter influência alguma sobre os resultados dos outros lançamentos. Desejamos agora construir um espaço de probabilidade para o experimento composto, consistindo de n repetições do nosso experimento, que incorpore as nossas crenças intuitivas.

Já que cada uma das n repetições pode resultar em sucesso ou fracasso, existem um total de 2^n resultados possíveis para o experimento composto. Pode-se representar esses resultados através de uma n-tupla $(x_1, x_2, \ldots, x_n)$, onde $x_i = 1$ se a i-ésima repetição resulta em sucesso e $x_i = 0$ caso contrário. Tomamos o conjunto Ω como sendo a coleção de todas as n-tuplas como esta. Toma-se o σ-álgebra $\mathscr{A}$ como sendo constituído de todos os subconjuntos de Ω.

Chegamos agora à alocação de uma medida de probabilidade. Para isso é necessário tão somente alocar probabilidades a 2^n conjuntos de um ponto $\{(x_1, x_2, \ldots, x_n)\}$. Suponha que a n-tupla $(x_1, \ldots, x_n)$ é tal que exatamente k dos x_i tem o valor 1, por simplicidade, digamos que $x_1 = x_2 = \cdots = x_k = 1$ e os outros x_i tenham o valor 0. Então se A_i representa o evento de que a i-ésima repetição resulta em sucesso, vemos que

$$\{(\underbrace{1, 1, \ldots, 1}_{k}, \underbrace{0, \ldots, 0}_{n-k})\} = A_1 \cap \cdots \cap A_k \cap A_{k+1}^c \cap \cdots \cap A_n^c.$$

De acordo com nossa intuição, os eventos $A_1, A_2, \ldots, A_k, A_{k+1}^c, \ldots, A_n^c$ devem ser mutuamente independentes e $P(A_i) = p$, $1 \leqslant i \leqslant n$. Assim devemos escolher P de tal forma que

$$P(\{(1, 1, \ldots, 1, 0, \ldots, 0)\}) = P(A_1) \cdots P(A_k) P(A_{k+1}^c) \cdots P(A_n^c)$$
$$= p^k(1 - p)^{n-k}.$$

Usando o mesmo raciocínio, vemos que se a n-tupla $(x_1, \ldots, x_n)$ é tal que exatamente k dos x_i tem o valor 1, P deve ser tal que

$$P(\{(x_1, \ldots, x_n)\}) = p^k(1 - p)^{n-k}.$$

Determinemos a seguir a probabilidade de que exatamente k das n repetições resultem em sucesso. Observe, cuidadosamente, que isto difere da probabilidade de que exatamente k repetições específicas, resultem em sucesso e que, as outras $n - k$ repetições, resultem em fracasso. Seja B_k o evento de que se obtém sucessos em exatamente k das n repetições. Uma vez que cada seqüência específica com k sucessos têm probabilidade $p^k(1 - p)^{n-k}$, o evento B_k tem probabilidade $P(B_k) = C(k, n)p^k(1 - p)^{n-k}$, onde $C(k, n)$ é o número de seqüências $(x_1, \ldots, x_n)$ nas quais exatamente k dos x_i tem valor 1. A determinação de $C(k, n)$ é um problema combinatório simples que será resolvido na Seção 2.4, onde mostraremos que

$$C(k, n) = \frac{n!}{k!\,(n - k)!}, \qquad 0 \leq k \leq n. \tag{20}$$

Lembre-se que $0! = 1$ e que

$$m! = m(m - 1) \cdots 1.$$

para qualquer número inteiro positivo m. Geralmente representa-se a quantidade $n!/k!\,(n - k)!$ por $\binom{n}{k}$ (coeficiente binomial). Assim

$$P(B_k) = \binom{n}{k} p^k(1 - p)^{n-k}. \tag{21}$$

Diversos problemas aplicados são modelados por provas independentes do tipo sucesso-fracasso. O problema discutido a seguir é típico.

Exemplo 9. Suponha que uma máquina produza parafusos dos quais 10% são defeituosos. Determine a probabilidade de que uma caixa com 3 parafusos contenha no máximo um parafuso defeituoso.

Para resolver o problema, supomos que a produção de parafusos constitui repetições independentes de uma prova do tipo sucesso-fracasso em que a produção de um parafuso defeituoso é um sucesso. Então a probabilidade de sucesso neste caso é 0,1. Seja B_0 o evento de que nenhum dos três parafusos é defeituoso e

B_1 o de que exatamente um dos três parafusos é defeituoso. Então $B_0 \cup B_1$ é o evento de que no máximo 1 parafuso é defeituoso. Como B_0 e B_1 são eventos disjuntos, segue-se que

$$\begin{aligned} P(B_0 \cup B_1) &= P(B_0) + P(B_1) = \\ &= \binom{3}{0}(0,1)^0(0,9)^3 + \binom{3}{1}(0,1)^1(0,9)^2 = \\ &= (0,9)^3 + 3(0,1)(0,9)^2 = \\ &= 0,972. \end{aligned}$$

Exercícios

1. Seja $(\Omega, \mathscr{A}, P)$ um espaço de probabilidade, onde $\mathscr{A}$ é σ-álgebra de todos os subconjuntos de Ω e P, uma medida de probabilidade que associa a probabilidade $p > 0$ cada conjunto de um ponto de Ω.
   (a) Mostre que Ω deve ter um número finito de pontos. *Sugestão*: mostre que Ω não pode ter mais de p^{-1} pontos.
   (b) Mostre que se n é o número de pontos em Ω, então p deve ser igual a n^{-1}.
2. Pode-se construir um modelo para um "spinner" aleatório tomando um espaço uniforme de probabilidade sobre a circunferência de um círculo de raio 1, de modo que a probabilidade de que o ponteiro do "spinner" pare sobre um arco de comprimento s é $s/2\pi$. Suponha que o círculo esteja dividido em 37 zonas numeradas de 1 a 37. Determine a probabilidade de que o ponteiro pare sobre uma zona par.
3. Considere um ponto escolhido ao acaso sobre um quadrado unitário. Determine a probabilidade de que o ponto esteja no triângulo limitado por $x = 0$, $y = 0$ e $x + y = 1$.
4. Seja P um ponto escolhido ao acaso sobre um disco unitário. Determine a probabilidade de que P esteja no setor angular de 0 a $\pi/4$ radianos.
5. No Exemplo 2 determine as probabilidades dos seguintes eventos:
   (a) Nenhuma desintegração ocorre antes do tempo 10.
   (b) Ocorre uma desintegração antes do tempo 2 ou uma entre os tempos 3 e 5.
6. Uma caixa contém 10 bolas numeradas de 1 a 10. Extrai-se ao acaso uma bola da caixa. Determine a probabilidade de que o número da bola seja 3, 4 ou 5.
7. Suponha que se lance um par de dados e que os 36 resultados possíveis são igualmente prováveis. Determine a probabilidade de que a soma dos números observados seja par.

8. Suponha que A e B sejam eventos tais que $P(A) = 2/5$, $P(B) = 2/5$ e $P(A \cup B) = 1/2$. Determine $P(A \cap B)$.

9. Se $P(A) = 1/3$, $P(A \cup B) = 1/2$, e $P(A \cap B) = 1/4$, determine $P(B)$.

10. Suponha que se escolha ao acaso um ponto sobre um quadrado unitário. Seja A o evento de que o ponto está no triângulo limitado por $y = 0$, $x = 1$ e $x = y$, e B o evento de que o ponto está no retângulo com vértices em $(0,0)$, $(1, 0)$, $(1, 1/2)$ e $(0, 1/2)$. Determine $P(A \cup B)$ e $P(A \cap B)$.

11. Uma caixa contém 10 bolas numeradas de 1 a 10. Seleciona-se uma bola ao acaso e a seguir seleciona-se, também ao acaso, uma segunda bola dentre as 9 restantes. Determine a probabilidade de que os números das bolas selecionadas diferem de duas ou mais unidades.

12. Dado que um ponto escolhido ao acaso sobre um quadrado unitário esteja no triângulo limitado por $x = 0$, $y = 0$ e $x + y = 1$, determine a probabilidade de que o mesmo esteja também no triângulo limitado por $y = 0, x = 1$ e $x = y$.

13. Suponha que temos quatro cofres, cada um com 2 gavetas. Os cofres 1 e 2, têm uma moeda de ouro em uma gaveta e uma de prata na outra. O cofre 3, tem duas moedas de ouro e, o cofre 4 tem duas de prata. Escolhe-se um cofre ao acaso, abre-se uma gaveta e encontra-se uma moeda de ouro. Determine a probabilidade de que a outra gaveta contenha
    (a) uma moeda de prata;
    (b) uma moeda de ouro.

14. Uma caixa contém 10 bolas das quais 6 são pretas e 4 são brancas. Remove-se três bolas sem observar suas cores. Determine a probabilidade de que uma quarta bola removida da caixa seja branca. Assuma que as 10 bolas são igualmente prováveis de serem removidas da caixa.

15. Para uma caixa de mesma composição que a do Exercício 14, determine a probabilidade de que todas as 3 bolas removidas sejam pretas, sabendo-se que pelo menos uma delas é preta.

16. Suponha que uma fábrica tem duas máquinas A e B, responsáveis, respectivamente, por 60% e 40% da produção total. A máquina A produz 3% de itens defeituosos, enquanto que a máquina B produz 5% de itens defeituosos. Determine a probabilidade de que um dado item defeituoso foi produzido pela máquina B.

17. Mostre por indução sobre n que a probabilidade de selecionar uma bola vermelha na n-ésima extração no esquema de Polya (Exemplo 7) é $r(b + r)^{-1}$. para qualquer n.

18. Um estudante se submete a um exame de múltipla escolha no qual cada questão tem 5 respostas possíveis das quais exatamente uma é correta. O estudante seleciona a resposta correta se ele sabe a resposta. Caso contrário, ele seleciona ao acaso uma resposta entre as 5 possíveis. Suponha que o estudante saiba a resposta de 70% das questões.

(a) Qual a probabilidade de que o estudante escolha a resposta correta para uma dada questão?

(b) Se o estudante escolhe a resposta correta para uma dada questão, qual a probabilidade de que ele sabia a resposta?

19. Suponha que se escolha ao acaso um ponto sobre um quadrado unitário. Sabendo-se que o ponto está no retângulo limitado por $y = 0$, $y = 1$, $x = 0$, e $x = 1/2$, qual é a probabilidade de que o ponto esteja no triângulo limitado por $y = 0$, $x = 1/2$ e $x + y = 1$?

20. Suponha que uma caixa contenha r bolas vermelhas e b bolas pretas. Extrai-se ao acaso uma bola da caixa e a seguir extrai-se, também ao acaso, uma segunda bola dentre as que ficaram na caixa. Determine a probabilidade de que

(a) ambas as bolas sejam vermelhas;

(b) a primeira bola seja vermelha e a segunda preta;

(c) a primeira bola seja preta e a segunda vermelha;

(d) ambas as bolas sejam pretas.

21. Uma caixa contém 10 bolas vermelhas e 5 pretas. Extrai-se uma bola da caixa. Se a bola é vermelha, ela é recolocada na caixa. Se é preta, além de recolocá-la na caixa, adiciona-se duas bolas à caixa. Determine a probabilidade de que uma segunda bola extraída da caixa seja

(a) vermelha; (b) preta.

22. Extrai-se duas bolas, com reposição da primeira, de uma caixa contendo 3 bolas brancas e 2 bolas pretas.

(a) Construa um espaço de amostra com pontos igualmente prováveis para este experimento.

(b) Determine a probabilidade de que as bolas extraídas sejam da mesma cor.

(c) Determine a probabilidade de que, pelo menos, uma das bolas extraídas seja branca.

23. Resolva o Exercício 22, caso a primeira bola não seja reposta na caixa.

24. Resolva o Exercício 22, construindo um espaço de amostra baseado em 4 pontos correspondendo a bola branca e bola preta para cada extração.

25. Caixa I contém 3 bolas brancas e 2 pretas, caixa II contém 2 bolas brancas e 1 preta e caixa III contém 1 bola branca e 3 pretas.

(a) Extrai-se uma bola de cada caixa. Determine a probabilidade de que todas as bolas sejam brancas.

(b) Seleciona-se uma caixa e dela extrai-se uma bola. Determine a probabilidade de que a bola extraída seja branca.

(c) Calcule em (b) a probabilidade de que a primeira caixa foi selecionada, dado que a bola extraída é branca.

26. Uma caixa contém 3 bolas brancas e 2 bolas pretas. Selecionam-se duas bolas sem reposição.

(a) Calcule a probabilidade de que a segunda bola seja preta, dado que a primeira é preta.

(b) Calcule a probabilidade de que a segunda bola seja da mesma cor da primeira.

(c) Calcule a probabilidade de que a primeira bola seja branca, dado que a segunda é branca.

27. Um colégio é composto de 70% de homens e 30% de mulheres. Sabe-se que 40% dos homens e 60% das mulheres são fumantes. Qual é a probabilidade de que um estudante que foi visto fumando seja homem?

28. Suponha que os automóveis têm igual probabilidade de serem produzidos na segunda, terça, quarta, quinta e sexta-feira. As porcentagens de automóveis amarelos produzidos nos diferentes dias da semana são: segunda – 4%; terça, quarta e quinta – 1%; e sexta – 2%. Se você compra um automóvel amarelo, qual a probabilidade de que o mesmo foi produzido numa segunda-feira?

29. Suponha que existisse um teste para câncer com a propriedade de que 90% das pessoas com câncer e 5% das pessoas sem câncer reagem positivamente. Admita que 1% dos pacientes de um hospital tem câncer. Qual a probabilidade de que um paciente escolhido ao acaso, que reage positivamente a esse teste, realmente tenha câncer?

30. No problema de três cofres discutidos no Exemplo 6, determine a probabilidade de que a segunda gaveta contenha uma moeda de prata, dado que a primeira continha uma moeda de ouro.

31. No esquema de urna de Polya (Exemplo 7), dado que a segunda bola é vermelha, determine a probabilidade de que

(a) a primeira bola era vermelha;

(b) a primeira bola era preta.

32. Suponha que se lança três moedas idênticas e perfeitamente equilibradas. Seja A_i o evento de observar cara na i-ésima moeda. Mostre que os eventos A_1, A_2 e A_3 são mutuamente independentes.

33. Suponha que as seis faces de um dado têm igual probabilidade de ocorrência e que sucessivos lançamentos do dado são independentes. Construa um espaço de probabilidade para o experimento composto de três lançamentos do dado.

34. Sejam A e B dois eventos independentes. Mostre que A e B^c, A^c e B e A^c e B^c são também independentes.

35. Seja $\Omega = \{1, 2, 3, 4\}$ e suponha que cada ponto tem probabilidade 1/4. Sendo $A = \{1, 2\}$, $B = \{1, 3\}$ e $C = \{1, 4\}$, mostre que os pares de eventos A e B, A e C e B e C são independentes.

36. Suponha que A, B e C são eventos mutuamente independentes e $P(A \cap B) \neq 0$. Mostre que $P(C \mid A \cap B) = P(C)$.

37. Experiência mostra que 20% das pessoas que fazem reservas de mesa num certo restaurante deixam de comparecer. Se o restaurante tem 50 mesas e aceita 52 reservas, qual a probabilidade de que seja capaz de acomodar todos os fregueses?

38. Um alvo circular de raio unitário está dividido em quatro zonas anelares com raios externos 1/4, 1/2, 3/4 e 1, respectivamente. Suponha que se dispare ao alvo 10 tiros independentemente e ao acaso.

    (a) Determine a probabilidade de que no máximo 2 tiros atinjam a zona limitada pelos círculos de raios 1/2 e 1.

    (b) Se 5 tiros atingem o disco de raio 1/2, determine a probabilidade de que pelo menos um atinja o disco de raio 1/4.

39. Uma máquina consiste de 4 componentes ligados em paralelo, de tal sorte que a máquina falha somente se todos os quatro componentes falham. Suponha que as falhas dos componentes são independentes umas das outras. Se os componentes tem probabilidades 0,1; 0,2; 0,3 e 0,4 de falhar quando a máquina é acionada, qual é a probabilidade de que a máquina funcione quando acionada?

40. Um certo componente de um motor de foguete falha 5% das vezes quando o motor é acionado. Para obter maior confiabilidade no funcionamento do motor, duplica-se esse componente n vezes. O motor então falha somente se todos os n componentes falham. Suponha que as falhas dos componentes sejam independentes uma das outras. Qual é o menor valor de n que pode ser usado para garantir que o motor funcione 99% das vezes?

41. Lança-se um dado simétrico três vezes. Sabendo-se que a face 1 ocorre pelo menos uma vez, qual é a probabilidade de que ela ocorra exatamente uma vez?

42. Existem 4 reis num baralho de 52 cartas. Extrai-se uma carta do baralho, registra-se o seu valor e a seguir repõe-se a carta extraída. Este procedimento é repetido 4 vezes. Determine a probabilidade de que existam exatamente 2 reis entre as cartas selecionadas, sabendo-se que entre elas existe pelo menos um rei.

43. Mostre que se A, B e C são eventos tais que $P(A \cap B \cap C) \neq 0$ e $P(C \mid A \cap B) = P(C \mid B)$, então $P(A \mid B \cap C) = P(A \mid B)$.

44. Um homem dispara 12 tiros independentes num alvo. Qual a probabilidade de que ele atinja o alvo pelo menos uma vez, se tem probabilidade 9/10 de atingir o alvo em qualquer tiro?

45. Lança-se um dado 12 vezes. Determine a probabilidade de obter

    (a) dois "seis";

    (b) no máximo dois "seis".

46. Suponha que a probabilidade de atingir um alvo é 1/4. Disparando-se oito tiros ao alvo, qual a probabilidade de que o alvo seja atingido pelo menos duas vezes?

47. No Exercício 44 qual a probabilidade de que o alvo seja atingido pelo menos duas vezes, sabendo-se que o mesmo foi atingido pelo menos uma vez?

2

ANÁLISE COMBINATÓRIA

Lembre-se da Seção 1.2 onde um espaço simétrico de probabilidade contendo s pontos, é o modelo usado para o experimento de selecionar ao acaso um ponto de um conjunto S contendo s pontos. Daqui para diante, quando falarmos em selecionar ao acaso um ponto de um conjunto finito S, estaremos dizendo que, a probabilidade associada a cada conjunto de um ponto é s^{-1}, e, portanto, que a probabilidade associada a um conjunto A contendo j pontos é j/s.

Seja $N(A)$ o número de pontos em A. Desde que $P(A) = N(A)/s$, o problema de determinar $P(A)$ é equivalente ao de determinar $N(A)$. O procedimento para obter $P(A)$ é contar o número de pontos em 4 e dividir pelo número total s de pontos. Entretanto, às vezes inverte-se o procedimento. Se de alguma forma conhecemos $P(A)$, podemos obter $N(A)$ pela fórmula $N(A) = sP(A)$. Este procedimento será usado muitas vezes na seqüência.

A determinação de $N(A)$ é fácil quando A tem apenas alguns pontos, pois neste caso podemos, simplesmente enumerar todos os pontos em A. Porém, mesmo que A tenha um número moderado de pontos, o método de enumeração direta torna-se impraticável, e assim algumas regras simples de contagem são desejáveis. Nosso propósito neste capítulo é apresentar uma discussão não-técnica e sistemática de métodos que são elementares porém de grande aplicabilidade. Este assunto tende a tornar-se difícil rapidamente, por isso limitaremos nosso tratamento àquelas partes maiores na teoria da probabilidade. As primeiras quatro seções deste capítulo contêm o material essencial, enquanto as quatro últimas seções contêm um material opcional e um pouco mais difícil.

2.1. AMOSTRAS ORDENADAS

Suponha que temos dois conjuntos S e T. Se S tem m pontos distintos $s_1, s_2, \ldots, s_m$ e T tem n pontos distintos $t_1, t_2, \ldots, t_n$, então o número de pares (s_i, t_j) que podem ser formados tomando um ponto do conjunto S e um segundo ponto do conjunto T é mn. Isto é claro de vez que qualquer elemento do conjunto S pode ser associado a qualquer um dos n elementos do conjunto T.

Exemplo 1. Se $S = \{1,2\}$ e $T = \{1,2,3\}$, então existem seis pares: (1,1), (1,2), (1,3), (2,1), (2,2), (2,3). Observe cuidadosamente que o par (1,2) é distinto do par (2,1).

De forma mais geral, suponha que temos n conjuntos $S_1, S_2, \ldots, S_n$ tendo $s_1, s_2 \ldots s_n$ pontos distintos, respectivamente. Então o número de n-tuplas $(x_1, x_2, \ldots, x_n)$ que podem ser formados, onde x_1 é um elemento de S_1, x_2 um elemento de $S_2, \ldots,$ e x_n um elemento de S_n, é $s_1 s_2 \ldots s_n$. Isto é uma extensão bastante óbvia do caso discutido acima para $n = 2$. (Uma demonstração formal de que o número de n-tuplas é $s_1 s_2 \ldots s_n$ pode ser feita por indução sobre n.)

Um importante caso especial ocorre quando cada conjunto S_i, $1 \leqslant i \leqslant n$, é o mesmo conjunto S tendo s pontos distintos. Existem então s^n n-tuplas $(x_1, x_2, \ldots, x_n)$, onde cada x_i é um ponto do conjunto S.

Exemplo 2. $S = \{1, 2\}$ e $n = 3$. Neste caso existem 8 n-tuplas: (1, 1, 1), (1, 1, 2), (1, 2, 1), (1, 2, 2), (2, 1, 1), (2, 1, 2), (2, 2, 1), (2, 2, 2).

O caso especial em que os conjuntos S_i, $1 \leqslant i \leqslant n$, são os mesmos, pode ser abordado de um ponto de vista diferente. Suponha que uma caixa contém s bolas distintas numeradas de 1 a s. Extrai-se uma bola da caixa, registra-se o seu número e a seguir repõe-se a bola na caixa. Repete-se o procedimento n vezes. Cada uma das n extrações produz um número de 1 a s. O resultado pode ser registrado através de uma n-tupla. $(x_1, x_2, \ldots, x_n)$, onde x_1 é o número da 1ª bola, x_2, o da 2ª etc. Existem ao todo s^n n-tuplas. Este procedimento chama-se *amostragem com reposição* de uma população de s objetos distintos. O resultado $(x_1, x_2, \ldots, x_n)$ chama-se amostra de tamanho n extraída com reposição de uma população de s objetos. Falamos de *amostragem aleatória com reposição*, se admitirmos que todas as s^n amostras possíveis têm a mesma probabilidade ou, na linguagem tradicional, são igualmente prováveis de ocorrer.

Exemplo 3. Lança-se n vezes uma moeda perfeitamente equilibrada. Determinar a probabilidade de obter pelo menos uma cara.

Presumivelmente a afirmação de que a moeda é perfeitamente equilibrada implica que a probabilidade de obter cara em qualquer lançamento é 1/2. Se assim for, e se supormos que lançar a moeda n vezes equivale extrair uma amostra aleatória de tamanho n de uma população de dois objetos $\{$ H, T $\}$, então cada um dos 2^n resultados possíveis é igualmente provável. Seja A o evento de obter pelo menos uma cara e A_i o de obter cara no i-ésimo lançamento. Então $A = \bigcup_{i=1}^{n} A_i$. Mas

$$P(A) = 1 - P(A^c) = 1 - P\left(\left(\bigcup_{i=1}^{n} A_i\right)^c\right)$$

$$= 1 - P\left(\bigcap_{i=1}^{n} A_i^c\right)$$

e $\bigcap_{i=1}^{n} A_i^c$ ocorre se e somente se, todos os n lançamentos produzirem coroas. Assim $P(\bigcap_{i=1}^{n} A_i^c) = 2^{-n}$, de modo que $P(A) = 1 - 2^{-n}$.

Seja S um conjunto contendo s objetos distintos. Selecionamos um objeto de S e registramos o objeto selecionado, mas suponha agora que não o repomos ao conjunto S. Se repetirmos este procedimento, faremos uma seleção dentre $(s - 1)$ objetos restantes. Suponha que o procedimento seja repetido mais $n - 1$ vezes, de modo que selecionamos n objetos ao todo. (Obviamente devemos ter $n \leqslant s$ neste caso). Novamente podemos registrar o resultado em forma de uma n-tupla $(x_1, x_2, \ldots, x_n)$, porém desta vez os números $x_1, x_2, \ldots, x_n$ devem ser distintos, não pode haver duplicações na nossa amostra. O primeiro objeto pode ser qualquer um dos s objetos, o segundo pode ser qualquer um dos $s - 1$ objetos restantes, o terceiro pode ser qualquer um dos $s - 2$ objetos restantes etc., de modo que existem ao todo $(s)_n = s(s - 1) \cdots (s - n + 1)$ diferentes resultados possíveis para o experimento. Refere-se a este procedimento como *amostragem sem reposição* n de uma população de s objetos distintos. Falamos de uma *amostra aleatória de tamanho* n *tomada sem reposição de uma população de* s *objetos* se supomos que cada um dos $(s)_n$ resultados é igualmente provável.

Representamos o produto $s(s - 1) \cdots (s - n + 1)$ por meio de símbolo $(s)_n$. Em particular $(s)_s = s(s - 1) \cdots 1 = s!$ Mas extrair uma amostra de tamanho s de uma população de s objetos diferentes equivale a escrever os números $1, 2, \ldots, s$ em alguma ordem. Assim $s!$ representa o número de ordenações diferentes (ou permutações) de s objetos.

Suponha que se extraia com reposição uma amostra aleatória de tamanho n de um conjunto de s objetos. Desejamos a probabilidade do evento A de que nenhum ponto ocorre duas vezes na amostra. Este problema pode ser resolvido facilmente. O número de amostras de tamanho n com reposição é s^n. Dentre estas s^n amostras aleatórias, o número daquelas nas quais nenhum ponto aparece duas vezes é igual ao número de amostras de tamanho n extraídas *sem reposição* dentre s objetos, isto é, s^n. Assim, já que todas as amostras são igualmente prováveis, vemos que a probabilidade desejada é

$$(1) \qquad \frac{(s)_n}{s^n} = \frac{s(s-1)\cdots(s-n+1)}{s^n}$$

$$= \left(1 - \frac{1}{s}\right)\left(1 - \frac{2}{s}\right)\cdots\left(1 - \frac{n-1}{s}\right).$$

Exemplo 4. Uma aplicação recente e um tanto surpreendente de (1) é no chamado problema do aniversário. Suponha que os aniversários das pessoas ocorram com igual probabilidade entre os 365 dias do ano. (Ignoramos aqui os anos bissextos e o fato de que as taxas de natalidade não são exatamente uniformes ao longo do ano). Determinar a probabilidade p de que em um grupo de n pessoas não existam duas com aniversários comuns.

Neste problema temos $s = 365$, e assim aplicando (1) vemos que

$$p = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right)\cdots\left(1 - \frac{n-1}{365}\right).$$

As conseqüências numéricas são bastantes inesperadas. Mesmo para s tão pequeno como 23, $p < 1/2$, e para $n = 56$, $p = 0{,}01$. Isto é, num grupo de 23 pessoas, a probabilidade de que pelo menos duas pessoas tenham o mesmo aniversário excede 1/2. Num grupo de 56 pessoas é quase certo que duas pessoas tenham o mesmo aniversário.

Se temos uma população de s objetos, existem s^n amostras de tamanho n que podem ser extraídas com reposição e $(s)_n$ amostras de tamanho n que podem ser extraídas sem reposição. Se s for grande em relação a n, existe pouca diferença entre esses dois métodos de amostragem. Realmente, vemos de (1) que para qualquer n fixo.

$$(2) \qquad \lim_{s\to\infty} \frac{(s)_n}{s^n} = \lim_{s\to\infty} \left(1 - \frac{1}{s}\right) \cdots \left(1 - \frac{n-1}{s}\right) = 1.$$

(Para estimativas mais precisas ver Exercício 12.)

2.2. Permutações

Suponha que temos n caixas distintas e n bolas distintas. O número total de maneiras de distribuir as n bolas nas caixas de modo que *cada caixa contenha exatamente uma bola é* $n!$ Dizer que essas bolas são distribuídas ao acaso nas n caixas com uma bola por caixa significa que associamos probabilidade $1/n!$ a cada uma dessas distribuições possíveis. Suponha que este seja o caso. Qual é a probabilidade de que uma bola específica, digamos bola i, esteja numa caixa específica, digamos caixa j? Se a bola i está na caixa j, isto deixa $(n - 1)$ bolas para serem distribuídas em $(n - 1)$ caixas de modo que exatamente uma bola esteja em cada caixa. Isto pode ser feito de $(n - 1)!$ maneiras, de modo que a probabilidade desejada é $(n - 1)!/n! = 1/n$.

Outra maneira de interpretar este resultado é a seguinte. Se permutamos entre si n objetos distintos, a probabilidade de que um objeto específico esteja numa posição específica é $1/n$. Realmente, aqui podemos identificar as posições com as caixas e os objetos com as bolas.

Pode-se estender facilmente as considerações acima de 1 para $k \geqslant 1$ objeto. Se permutamos entre si n objetos, a probabilidade de que k objetos específicos estejam em k posições específicas é $(n - k)!/n!$. Deixamos a demonstração desse fato para o leitor.

Problemas envolvendo permutações aleatórias, tomam uma variedade de formas quando enunciados literalmente. Apresentamos a seguir dois exemplos:

(a) Mistura-se um baralho de cartas numeradas de 1 a n e extrai-se as cartas uma a uma. Para um dado i, qual é a probabilidade de que a i-ésima carta extraída é a de número i?

(b) Suponha que dez casais comparecem a uma festa. Então forma-se ao acaso pares de rapazes e moças. Qual é a probabilidade de que exatamente k rapazes específicos fiquem com as respectivas namoradas?

Um problema mais sofisticado envolvendo permutações aleatórias, é o de determinar a probabilidade de que ocorram exatamente k "encontros". Para usar o nosso exemplo visual de distribuir bolas em caixas, o problema é determinar a probabilidade de que a bola i esteja na caixa i para exatamente k valores diferentes de i.

Pode-se resolver o problema do encontro de várias maneiras. Adiamos a discussão deste problema para a Seção 2.6.

2.3. Combinações (amostras desordenadas)

Uma mão de pôquer consiste de cinco cartas extraídas de um baralho de 52 cartas. De acordo com a discussão acima, existiriam $(52)_5$ mãos de pôquer. Entretanto, para chegar a esta contagem diferentes ordenações das mesmas cinco cartas são consideradas mãos diferentes. Isto é, a mão 2, 3, 4, 5, 6 de espadas nesta ordem é considerada diferente da mão 2, 4, 3, 5, 6 de espadas nesta ordem. Do ponto de vista do jogo, estas mãos são as mesmas. De fato, todas as 5! permutações das mesmas cinco cartas são equivalentes. Das $(52)_5$ mãos possíveis, exatamente 5! delas são simplesmente permutações dessas mesmas cinco cartas. Similarmente, para qualquer conjunto de cinco cartas existem 5! diferentes permutações. Assim, o número total de mãos de pôquer, *sem considerar a ordem em que as cartas aparecem,* é $(52)_5/5!$ Nesta contagem considera-se duas mãos diferentes se, e somente se, elas diferem como conjunto de objetos, isto é, se elas têm pelo menos um elemento diferente. Por exemplo, entre as $(52)_5/5!$ mãos de pôquer, as mãos (2, 3, 4, 5, 6) de espadas e (3, 2, 4, 5, 6) de espadas são as mesmas, porém as mãos (2, 3, 4, 5, 7) de espadas e (2, 3, 4, 5, 6) de espadas são diferentes.

De uma forma mais geral, suponha que temos um conjunto S de s objetos distintos. Então, como foi explicado anteriormente, existem $(s)_r$ amostras distintas de tamanho r que podem ser extraídas sem reposição de S. Cada subconjunto distinto $\{x_1, ..., x_r\}$ de r objetos de S pode ser ordenado (rearranjado) de $r!$ maneiras diferentes. Se decidimos ignorar a ordem em que os objetos aparecem na amostra, essas $r!$ reordenações de $x_1, \ldots, x_r$ serão consideradas as mesmas. Existem portanto, $(s)_r/r!$ *diferentes amostras de tamanho r que podem ser extraídas, sem reposição e sem consideração de ordem, de um conjunto de s objetos distintos.*

Geralmente representa-se a quantidade $(s)_r/r!$ por meio do símbolo do coeficiente binomial

$$\frac{(s)_r}{r!} = \binom{s}{r}.$$

Observe que para $r = 0, 1, 2, \ldots, s$

$$\binom{s}{r} = \frac{(s)_r}{r!} = \frac{s!}{r!(s-r)!}.$$

Indicamos aqui para uso futuro que $\binom{a}{r}$ é bem definido para qualquer número real a e qualquer número inteiro não-negativo r através de

$$(3) \qquad \binom{a}{r} = \frac{(a)_r}{r!} = \frac{a(a-1)\cdots(a-r+1)}{r!},$$

onde $0!$ e $(a)_0$ são ambos definidos como sendo 1.

Exemplo 5.

$$\binom{-\pi}{3} = \frac{(-\pi)(-\pi-1)(-\pi-2)}{3!}$$
$$= -\frac{\pi(\pi+1)(\pi+2)}{3!}.$$

Observe que se a é um número inteiro positivo, então $\binom{a}{r} = 0$ para $r > a$. Convencionamos que $\binom{a}{r} = 0$ se r é um número inteiro negativo. Então $\binom{a}{r}$ é definido para todo real a e todo inteiro r.

Como foi observado anteriormente, quando s é um inteiro positivo e r é um inteiro não-negativo, é interessante pensar em $\binom{s}{r}$ como o número de maneiras em que podemos extrair sem reposição uma amostra de tamanho r de uma população de s objetos distintos sem consideração da ordem em que esses objetos são escolhidos.

Exemplo 6. Considere o conjunto de números $\{1, 2, \ldots, n\}$. Então, se $1 \leqslant r \leqslant n$, existem exatamente $\binom{n}{r}$ escolhas de números $i_1, i_2, \ldots, i_r$ tais que $1 \leqslant i_1 < i_2 < \cdots < i_r \leqslant n$. Realmente, cada uma das $(n)_r$ escolhas de r números distintos de 1 a n tem $r!$ reordenações das quais exatamente uma satisfaz a condição desejada. Assim, o número de escolhas distintas de números que satisfazem a condição é igual ao número de distintos subconjuntos de tamanho r que se pode extrair do conjunto $\{1, 2, \ldots, n\}$.

Exemplo 7. Membros de comissão. O departamento de matemática consiste de 25 professores titulares, 15 professores adjuntos e 35 professores assistentes. Uma comissão de 6 membros é selecionada ao acaso do corpo docente do departamento. Determine a probabilidade de que todos os membros da comissão sejam professores assistentes.

Existem ao todo 75 professores. De 75 uma comissão de 6 pode ser escolhida de $\binom{75}{6}$ maneiras. Como são 35 professores assistentes, existem $\binom{35}{6}$ maneiras de formar uma comissão como apenas professores assistentes. Assim a probabilidade desejada é $\binom{35}{6} / \binom{75}{6}$. Cálculos conduzem ao valor de 0,01, portanto os professores titulares e adjuntos não precisam preocupar-se indevidamente com a possibilidade de não ter representação na comissão.

Exemplo 8. Considere uma mão de pôquer de cinco cartas. Determine a probabilidade de obter um "four" (isto é, quatro cartas de mesmo valor) supondo que as cinco cartas são escolhidas ao acaso.

Podemos resolver o problema da seguinte forma.

Existem $\binom{52}{5}$ mãos diferentes, que deverão ser igualmente prováveis. Assim Ω terá $\binom{52}{5}$ pontos. Para que o evento desejado ocorra, devemos ter quatro cartas de mesmo valor. Existem 13 escolhas diferentes para o valor das cartas que irão compor o "four", a saber, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K e A. Para cada uma dessas escolhas (que determine quatro das cinco cartas da mão desejada) existem 48 cartas dentre as quais podemos escolher a 5ª carta. Já que cada uma das 13 escolhas do "four" pode ser associada com cada uma das 48 escolhas da 5ª carta, existem ao todo (13)(48) maneiras possíveis de obter uma mão de pôquer com quatro das cinco cartas iguais. A probabilidade desejada é portanto

$$\frac{(13)(48)}{\binom{52}{5}} \approx 2{,}40 \times 10^{-4}.$$

Exemplo 9. Suponha que se distribui n bolas em n caixas de tal forma que os n^n arranjos possíveis são igualmente prováveis. Determine a probabilidade de que somente a caixa 1 esteja vazia.

Neste caso o espaço de probabilidade consiste de n^n pontos igualmente prováveis. Seja A o evento de que apenas a caixa 1 esteja vazia. Isto pode acontecer somente se as n bolas estão nas $n - 1$ caixas restantes de tal forma que nenhuma caixa esteja vazia. Assim, exatamente uma dessas $(n - 1)$ caixas deve conter duas bolas, e as $(n - 2)$ caixas restantes devem ter exatamente uma bola cada. Seja B_j o evento de que a caixa j, $j = 2, 3, \ldots n$, tenha duas bolas, caixa 1 esteja vazia e as $(n - 2)$ caixas restantes tenham exatamente uma bola cada. Então os eventos B_j são disjuntos e $A = \bigcup_{j=2}^{n} B_j$. Para determinar $P(B_j)$ observe que as duas bolas colocadas na caixa j podem ser escolhidas dentre n bolas de $\binom{n}{2}$ maneiras. As $(n - 2)$ bolas podem ser rearranjadas de $(n - 2)!$ maneiras nas $(n - 2)$ caixas restantes. Assim, o número de maneiras distintas em que podemos colocar duas bolas na caixa j, deixar a caixa 1 vazia e colocar exatamente uma bola em cada uma das caixas restantes é $\binom{n}{2}(n - 2)!$. Portanto

$$P(B_j) = \frac{\binom{n}{2}(n - 2)!}{n^n}$$

e conseqüentemente

$$P(A) = \frac{(n - 1)\binom{n}{2}(n - 2)!}{n^n} = \frac{\binom{n}{2}(n - 1)!}{n^n}.$$

2.4. PARTIÇÕES

Uma grande variedade de problemas combinatórios que envolvem amostras desordenadas são do seguinte tipo. Uma caixa contém n bolas vermelhas e b bolas brancas. Extrai-se sem reposição da caixa uma amostra aleatória de tamanho n. Qual é a probabilidade de que a amostra contenha exatamente k bolas vermelhas (e portanto $n - k$ bolas pretas?)

Para resolver o problema, argumentamos como segue. Estamos interessados apenas no número total de bolas vermelhas e bolas pretas e não na ordem em que elas são extraídas. Isto é, estamos lidando com amostragem sem reposição e sem consideração da ordem. Podemos, portanto, tomar o nosso espaço de amostra como sendo a coleção de $\binom{b+r}{n}$ amostras de tamanho n que podem ser extraídas dessa maneira das $r + b$ bolas da população. Associamos a probabilidade $\binom{b+r}{n}^{-1}$ a cada uma dessas $\binom{b+r}{n}$ amostras. Devemos determinar a seguir o número de maneiras em que se pode extrair uma amostra de tamanho n, de modo que contenha exatamente k bolas vermelhas. De r bolas vermelhas pode-se extrair k bolas vermelhas de $\binom{r}{k}$ maneiras, sem considerar a ordem, e de b bolas pretas pode-se escolher $n - k$ bolas pretas de $\binom{b}{n-k}$ maneiras, sem considerar a ordem. Como podemos associar cada escolha de k bolas vermelhas com cada escolha de $n - k$ bolas pretas, existem um total de $\binom{r}{k}\binom{b}{n-k}$ escolhas possíveis. Assim a probabilidade desejada é

$$\frac{\binom{r}{k}\binom{b}{n-k}}{\binom{r+b}{n}}.$$

A essência deste tipo de problema é que a população (neste caso as bolas) é particionada em duas classes (bolas vermelhas e pretas). Toma-se uma amostra aleatória de um certo tamanho e deseja-se determinar a probabilidade de que a amostra contenha números específicos de itens das duas classes.

Em alguns problemas deste tipo as duas classes não são definidas explicitamente, mas podem ser identificadas a partir dos enunciados.

Exemplo 10. Uma mão de pôquer consiste de cinco cartas extraídas de um baralho comum de 52 cartas. Obtenha a probabilidade de que a mão tenha exatamente dois reis.

Para resolver o problema, observe que existem $\binom{52}{5}$ mãos de pôquer. Num baralho existem 4 reis e outras 48 cartas. Isto particiona as cartas em duas classes, reis e não-reis, tendo 4 e 48 cartas, respectivamente. A mão de pôquer é uma amostra de tamanho 5 extraída sem reposição e sem consideração de ordem da população de 52 cartas. O problema é assim o de determinar a probabilidade de

que a amostra contenha 2 membros da primeira classe e 3 da segunda. Portanto a probabilidade desejada é

$$\frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}} \approx 3{,}99 \times 10^{-2}.$$

Exemplo 11. Um baralho de cartas tem 4 naipes de 13 cartas: paus, ouros, copas e espadas.

(a) Qual é a probabilidade de que em uma mão de 5 cartas exatamente 3 sejam paus?

(b) Qual é a probabilidade de que em uma mão de 5 cartas exatamente 3 sejam do mesmo naipe?

Para resolver (a) observe que as condições do problema dividem o baralho de 52 cartas em duas classes. A primeira é a de "paus" com 13 membros, e a classe dois é a de "não-paus" com 39 membros. As 5 cartas constituem uma amostra de tamanho 5 de uma população de 52 cartas, e o problema requer que 3 das 5 cartas sejam da classe 1. Assim, a probabilidade desejada é

$$p = \frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}} \approx 8{,}15 \times 10^{-2}.$$

Para resolver (b), seja A_1 o evento de que exatamente três cartas sejam paus, A_2 o de que exatamente três sejam ouros, A_3 o de que exatamente três sejam copas e A_4 o de que exatamente três sejam espadas. Então, como existem apenas 5 cartas na mão, os eventos A_1, A_2, A_3 e A_4 são mutuamente disjuntos. Sua união $A_1 \cup A_2 \cup A_3 \cup A_4$ é o evento de que exatamente 3 das 5 cartas sejam do mesmo naipe. Assim a probabilidade desejada é $4p$.

Exemplo 12. Considere novamente uma mão de pôquer de 5 cartas. Qual é a probabilidade de que ela seja um "full house" (isto é, um par de cartas de igual valor e uma trinca de cartas de igual valor), supondo que as cartas são extraídas do baralho ao acaso?

Para resolver o problema, observamos novamente que existem $\binom{52}{5}$ mãos equiprováveis de pôquer. Dentre elas devemos agora determinar o número de maneiras em que podemos ter um par e uma trinca. Considere o número de maneiras em que podemos escolher uma *trinca particular*, digamos 3 ases, e um *par particular*, digamos 2 reis. A trinca tem 3 cartas que serão escolhidas sem consideração de ordem dentre 4 ases, e isso pode ser feito de $\binom{4}{3}$ maneiras. O par tem 2 cartas que serão escolhidas sem consideração de ordem dentre 4 reis. Isso pode ser feito de $\binom{4}{2}$ maneiras. Então o número total de maneiras de se obter uma mão com

uma trinca de ases e um par de reis é $\binom{4}{3}\binom{4}{2}$. Assim, a probabilidade de obter uma mão com uma trinca de ases e um par de reis é $\binom{4}{3}\binom{4}{2}/\binom{52}{5} = p$. Naturalmente esta probabilidade seria a mesma para qualquer par específico e qualquer trinca específica. Mas o valor das cartas da trinca pode ser qualquer um dos 13, e o valor das cartas do par, pode ser qualquer um dos 12 restantes. Como cada um dos 13 valores da trinca pode ser associado a cada um dos 12 valores do par, existem (13)(12) escolhas desse tipo. Cada uma dessas escolhas constituem um evento disjunto tendo probabilidade p, de modo que a probabilidade desejada é

$$(13)(12)p = \frac{(13)(12)(4)(6)}{\binom{52}{5}} \approx 1{,}44 \times 10^{-3}.$$

Exemplo 13. Qual é a probabilidade de obter exatamente dois pares? Aqui uma mão como $(2, 2, 2, 2, x)$ não conta como dois pares mas como uma quadra.

Para resolver o problema observamos que se a mão tem dois pares, então duas cartas tem o mesmo valor x, duas tem o mesmo valor $y \neq x$ e a quinta carta tem um valor diferente de x e y. Existem 13 valores diferentes. Os valores dos pares podem ser escolhidos de $\binom{13}{2}$ maneiras. A outra carta pode ter qualquer um dos 11 valores restantes. As duas cartas de valor x podem ser escolhidas dentre 4 cartas deste valor de $\binom{4}{2}$ maneiras, o mesmo valendo para as duas cartas de valor y. A última carta de valor z pode ser escolhida de $\binom{4}{1} = 4$ maneiras dentre as quatro com esse valor. Assim o número total de escolhas é $\binom{13}{2}(11)\binom{4}{2}\binom{4}{2}(4)$, e portanto a probabilidade desejada é

$$\frac{\binom{13}{2}(11)(6)(6)(4)}{\binom{52}{5}} \approx 4{,}75 \times 10^{-2}.$$

Em alguns problemas envolvendo partições as classes são imaginadas como no exemplo a seguir.

Exemplo 14. Suponha que temos uma caixa contendo r bolas numeradas de 1 a r. Torna-se uma amostra aleatória sem reposição de tamanho n e registra-se os números das bolas. Repõe-se as bolas a caixa e torna-se uma segunda amostra aleatória sem reposição de tamanho m. Determine a probabilidade de que duas amostras tenham exatamente k bolas em comum.

Para resolver este problema podemos argumentar como segue. O efeito da primeira amostra é participar as bolas em duas classes: as n bolas selecionadas e as $r - n$ não-selecionadas (podemos imaginar que as n bolas da primeira amostra são pintadas de vermelho antes de serem repostas nas caixas). O problema é então

determinar a probabilidade de que a amostra de tamanho n contenha exatamente k bolas da primeira classe, de modo que a probabilidade desejada é

$$\frac{\binom{n}{k}\binom{r-n}{m-k}}{\binom{r}{m}}.$$

Se o argumento fosse conduzido de forma inversa e pensássemos na segunda amostra fazendo a partição obteríamos a probabilidade

$$\frac{\binom{m}{k}\binom{r-m}{n-k}}{\binom{r}{n}}.$$

Deixamos como exercício mostrar que essas duas probabilidades são iguais.

Podemos estender facilmente a nossa consideração de dividir uma população em duas classes para $m \geqslant 2$ classes. Suponha que temos um conjunto de r objetos tais que cada objeto é de um dos m tipos possíveis. A população consiste de r_1 objetos do tipo 1, r_2 objetos do tipo 2, . . . , r_m objetos do tipo m, onde $r_1 + r_2 + \cdots + r_m = r$. Extraindo-se uma amostra aleatória sem reposição de tamanho n da população desses r objetos, qual é a probabilidade de que a amostra contenha exatamente k_1, objetos do tipo 1, . . . , k_m objetos do tipo m onde $k_1 + \cdots + k_m = n$?

Uma vez mais o espaço de probabilidade é a coleção das $\binom{r}{n}$ amostras igualmente prováveis de tamanho n que se pode extrair sem reposição e sem consideração de ordem da população de r objetos. Pode-se escolher os k_i objetos do tipo i na amostra dos r_i objetos deste tipo na população sem consideração de ordem de $\binom{r_i}{k_i}$ maneiras. Assim, a probabilidade de escolher a amostra com a composição especificada é

$$\frac{\binom{r_1}{k_1}\binom{r_2}{k_2}\cdots\binom{r_m}{k_m}}{\binom{r}{n}}.$$

Exemplo 15. Em uma mão de 13 cartas de um baralho comum, obtenha a probabilidade de ocorrer exatamente 3 paus, 4 ouros, 4 copas e 2 espadas.

Neste problema $r = 52$ e $n = 13$. Se a classe 1 é constituída de paus, classe 2 de ouros, classe 3 de copas e classe 4 de espadas, então $m = 4, k_1 = 3, k_2 = 4, k_3 = 4$, e $k_4 = 2$, de modo que a probabilidade desejada é

$$\frac{\binom{13}{3}\binom{13}{4}\binom{13}{4}\binom{13}{2}}{\binom{52}{13}}$$

Exemplo 16. Problema de comissão. No problema da comissão discutido anteriormente, obtenha a probabilidade de que uma comissão de 6 membros seja composta de 2 professores titulares, 3 professores adjuntos e 1 professor assistente.

Usando o método acima, obtemos a resposta como sendo

$$\frac{\binom{25}{2}\binom{15}{3}\binom{35}{1}}{\binom{75}{6}}.$$

2.5. UNIÃO DE EVENTOS*

Considere novamente a permutação aleatória de n objetos distintos. Dizemos que ocorre um encontro na i-ésima posição se o i-ésimo objeto ocupa a i-ésima posição. Seja A_i o evento de que ocorre um encontro na posição i. Então $A = \bigcup_{i=1}^{n} A_i$ é o evento de que ocorre pelo menos um encontro. Podemos determinar $P(\bigcup_{i=1}^{n} A_i)$ para $n = 2$ através da Equação (10) do Capítulo 1 que estabelece que

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

É possível usar esta fórmula para obter uma fórmula semelhante para $n = 3$. Sejam A_1, A_2 e A_3 três eventos e seja $B = A_1 \cup A_2$. Então

$$P(A_1 \cup A_2 \cup A_3) = P(B \cup A_3) = P(B) + P(A_3) - P(B \cap A_3).$$

Mas

(4) $$P(B) = P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

como $B \cap A_3 = (A_1 \cup A_2) \cup A_3 = (A_1 \cap A_3) \cup (A_2 \cap A_3)$, segue-se que

(5) $$P(B \cap A_3) = P(A_1 \cap A_3) + P(A_2 \cap A_3) - P(A_1 \cap A_2 \cap A_3).$$

Substituindo (4) e (5) na expressão de $P(A_1 \cup A_2 \cup A_3)$, vemos que

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= [P(A_1) + P(A_2) - P(A_1 \cap A_2)] + P(A_3) \\ &\quad - [P(A_1 \cap A_3) + P(A_2 \cap A_3) - P(A_1 \cap A_2 \cap A_3)] \\ &= [P(A_1) + P(A_2) + P(A_3)] \\ &\quad - [P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3)] \\ &\quad + P(A_1 \cap A_2 \cap A_3). \end{aligned}$$

Para expressar esta fórmula de maneira mais conveniente, fazemos

$$S_1 = P(A_1) + P(A_2) + P(A_3),$$

$$S_2 = P(A_1 \cap A_2) + P(A_1 \cap A_3) + P(A_2 \cap A_3),$$

e

$$S_3 = P(A_1 \cap A_2 \cap A_3).$$

Então

(6) $$P(A_1 \cup A_2 \cup A_3) = S_1 - S_2 + S_3.$$

Existe uma generalização de (6) que é válida para todo número inteiro positivo n. Considere os eventos $A_1, A_2, \ldots, A_n$. Defina n números S_r, $1 \leqslant r \leqslant n$, através de

$$S_r = \sum_{1 \leq i_1 < \cdots < i_r \leq n} P(A_{i_1} \cap \cdots \cap A_{i_r}).$$

Então, em particular

$$S_1 = P(A_1) + \cdots + P(A_n),$$

$$S_2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_i \cap A_j),$$

e

$$S_n = P(A_1 \cap \cdots \cap A_n).$$

A fórmula desejada para $P(\bigcup_{i=1}^n A_i)$ é dada por

(7) $$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{r=1}^n (-1)^{r-1} S_r$$
$$= S_1 - S_2 + \cdots + (-1)^{n-1} S_n.$$

O leitor pode facilmente verificar que esta fórmula está de acordo com (6) se $n = 3$ e com a Equação 10 do Capítulo 1 se $n = 2$. A demonstração de (7) se faz por indução, sendo análoga a de (6). Omitiremos os detalhes da demonstração.

A soma S_1 tem n termos, a soma S_2 tem $\binom{n}{2}$ termos, e em geral a soma S_r tem $\binom{n}{r}$ termos. Isto pode ser verificado observando que a r-ésima soma é simplesmente a soma dos números $P(A_{i_1} \cap \cdots \cap A_{i_r})$ sobre todos os valores dos índices $i_1, i_2, \ldots, i_r$ tais que $i_1 < i_2 < \cdots < i_r$. Os índices assumem valores entre 1 e n. Assim o número de diferentes valores que estes índices podem assumir é igual ao número de maneiras diferentes em que podemos selecionar r números distintos dentre n números, sem reposição e sem consideração de ordem.

2.6. PROBLEMAS DE ENCONTRO

Podemos agora resolver facilmente o problema do número de encontros. Seja A_i o evento de que ocorre um encontro na posição i e seja p_n a probabilidade de que não ocorra nenhum encontro. Para determinar $1 - p_n = P(\bigcup_{i=1}^n A_i)$, precisamos determinar $P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r})$ onde $i_1, i_2, \ldots, i_r$ são r números distintos $\{1, 2, \ldots, n\}$. Mas esta probabilidade é simplesmente a probabilidade de ocorrer um encontro em cada uma das posições $i_1, i_2, \ldots, i_r$, e já determinamos

esta probabilidade como sendo $(n - r)!/n!$. Como a r-ésima soma S_r tem exatamente $\binom{n}{r}$ termos, vimos que

$$P(A_1 \cup \cdots \cup A_n) = \sum_{r=1}^{n} \binom{n}{r} \frac{(n-r)!}{n!} (-1)^{r-1}$$

$$= \sum_{r=1}^{n} (-1)^{r-1} \frac{n!}{r!(n-r)!} \frac{(n-r)!}{n!}$$

$$= \sum_{r=1}^{n} \frac{(-1)^{r-1}}{r!};$$

isto é,

$$(8) \qquad (1 - p_n) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{(-1)^{n-1}}{n!}.$$

Usando (8), vemos que a probabilidade p_n de que não ocorra encontros é

$$(9) \qquad p_n = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

Vemos que o segundo membro de (9) é simplesmente os primeiros $n + 1$ termos da expansão de e^{-1} em série de Taylor. Portanto, podemos aproximar p_n por e^{-1} e obter $1 - e^{-1} = 0{,}6321\ldots$ como uma aproximação de $(1 - p_n)$. Verifica-se que esta aproximação é *extremamente boa,* mesmo para valores pequenos de n. Na tabela abaixo apresentamos valores de $(1 - p_n)$ para diversos valores de n.

n	3	4	5	6
$1 - p_n$	0,6667	0,6250	0,6333	0,6320

Temos assim, o extraordinário resultado de que a probabilidade de ocorrer pelo menos um encontro entre n objetos permutados aleatoriamente é praticamente independente de n.

O problema de encontros pode ser reformulado de várias maneiras diferentes. Uma das mais famosas é a seguinte: Toma-se dois baralhos equivalentes de cartas e forma-se pares com uma carta de cada baralho. Qual é a probabilidade de ocorrer pelo menos um encontro?

Para resolver o problema precisamos observar apenas que se pode usar o primeiro baralho para determinar as posições (caixas). Sem perda de generalidade podemos supor que as cartas do primeiro baralho estão arranjadas na ordem $1, 2, \ldots, n$. Associa-se então as cartas do segundo baralho (bolas) às posições determinadas pelo primeiro baralho. Um encontro ocorre na posição i se, e somente se, a i-ésima carta extraída é a de número i.

Agora que sabemos como determinar a probabilidade p_n de nenhum encontro, podemos obter facilmente a probabilidade $\beta_n(r)$ de que haja exatamente r encontros. Para resolver o problema, determinamos inicialmente a probabilidade de que haja exatamente r encontros e que estes ocorram nas primeiras r posições. Isto pode acontecer somente se não houver encontros nas $(n - r)$ posições restantes. A probabilidade de que não haja encontros em j objetos permutados aleatoriamente é p_j. Portanto, $j!\ p_j$ é o número de maneiras em que se pode permutar j objetos entre si de modo que haja encontros (Por quê?). Já que existe apenas uma maneira de ter r encontros nas primeiras r posições o número de maneiras em que podemos ter r encontros nas primeiras r posições e não ter nenhum encontro nas $(n - r)$ posições restantes é $(n-r)!\ P_{n-r}$. Assim, a probabilidade desejada é

$$\alpha_r = \frac{(n - r)!}{n!} p_{n-r}.$$

A probabilidade de que haja exatamente r encontros e que estes ocorram em r posições específicas quaisquer é a mesma para todas as especificações das posições, isto é, α_r.

Para resolver o problema de que haja exatamente r encontros, tudo que é necessário perceber agora é que os eventos "exatamente r encontros ocorrendo nas posições $i_1, i_2, \ldots, i_r$" são eventos disjuntos para as diversas escolhas de $i_1, i_2, \ldots, i_r$. O número de tais escolhas é $\binom{n}{r}$. Assim, a probabilidade desejada é $\binom{n}{r}\alpha_r$. Portanto, se $\beta_n(r)$ é a probabilidade de que ocorra exatamente r encontros entre n objetos permutados aleatoriamente, obtemos

$$\text{(10)} \qquad \beta_n(r) = \frac{n!\,\alpha_r}{r!\,(n - r)!}$$

$$= \frac{n!}{r!\,(n - r)!} \frac{(n - r)!\, p_{n-r}}{n!}$$

$$= \frac{p_{n-r}}{r!}$$

$$= \frac{1}{r!}\left[1 - 1 + \frac{1}{2!} + \cdots + \frac{(-1)^{n-r}}{(n - r)!}\right].$$

Usando o fato de que p_{n-r} é aproximadamente igual a e^{-1} (aproximação que é bastante boa mesmo para $n - r$ moderadamente grande) obtemos

$$\text{(11)} \qquad \beta_n(r) \approx \frac{e^{-1}}{r!}.$$

Como ilustração final dessas idéias, determinaremos a probabilidade de que ocorra um encontro na posição j, dado que ocorrem exatamente r encontros.

Para resolver este problema, seja A_j o evento de que ocorre um encontro na posição j e B_r o evento de que ocorram exatamente r encontros. A probabilidade desejada é $P(A_j/B_r)$. De (10) temos $P(B_r) = p_{n-r}/r!$, de modo que precisamos determinar $P(A_j \cap B_r)$. Mas o evento $A_j \cap B_r$ ocorre se, e somente se, existe um encontro na j-ésima posição e existem exatamente $(r - 1)$ encontros nas $(n - 1)$ posições restantes. O número de maneiras em que podemos ter exatamente $(r - 1)$ encontros em $(n - 1)$ posições é $(n - 1)!\ \beta_{n-1}(r - 1)$. Assim

$$P(A_j \cap B_r) = \frac{(n-1)!\ \beta_{n-1}(r-1)}{n!}$$

$$= \frac{p_{n-r}}{(r-1)!\ n}.$$

Portanto

$$P(A_j \mid B_r) = \frac{p_{n-r}}{n(r-1)!}\ \frac{r!}{p_{n-r}} = \frac{r}{n}.$$

2.7. PROBLEMAS DE OCUPAÇÃO

Uma grande variedade de problemas combinatórios de probabilidade são equivalentes ao problema de distribuir n bolas distintas em r caixas distintas. Como qualquer uma das n bolas pode ocupar qualquer uma das r caixas, existem ao todo r^n maneiras diferentes de distribuir as bolas nas caixas. Supondo que a distribuição é feita ao acaso, cada uma dessas maneiras tem probabilidade r^{-n}. O espaço de probabilidade Ω para este experimento tem portanto r^n pontos igualmente prováveis. Nos problemas envolvendo esta distribuição de bolas, impomos diversas condições sobre a ocupação das caixas e perguntamos qual a probabilidade de que ocorra a situação especificada. Como primeiro exemplo considere o problema seguinte.

Distribuindo-se aleatoriamente n bolas em r caixas, qual é a probabilidade de que nenhuma caixa contenha mais de uma bola?

Para resolver o problema, observe antes de mais nada que a probabilidade desejada é 0 se $n > r$, por isso suponha que $n \leq r$. Então (pensando na distribuição das bolas uma a uma) a primeira bola pode ir para qualquer uma das r caixas, a segunda para qualquer uma das $(r - 1)$ caixas restantes etc., de modo que existem ao todo $(r)_n$ maneiras diferentes. A probabilidade desejada é então $(r)_n/r^n$.

Esta probabilidade é exatamente a mesma de extrair com reposição uma amostra de tamanho n de uma população de r objetos e obter uma amostra em que todos os objetos são distintos. Observe também que r^n é o número de amostras de tamanho n de uma população de r objetos distintos. Isto não é acidental. Extrair aleatoriamente n objetos com reposição é formalmente o mesmo que

distribuir aleatoriamente n bolas em r caixas. Isto pode ser constatado simplesmente pensando na distribuição das bolas nas caixas da forma seguinte. Extraímos primeiro uma amostra de tamanho n de um conjunto de r objetos, e se o i-ésimo elemento de amostra for o j-ésimo objeto, colocamos a bola i na caixa j. Às vezes é útil pensar na amostragem aleatória com reposição desta maneira, isto é, como distribuição aleatória de bolas em caixas (veja o problema de cupom ao final do capítulo).

Considere a distribuição aleatória de n bolas em r caixas. Qual é a probabilidade de que uma bola específica, digamos bola j, esteja em uma caixa específica, digamos caixa i? Se a bola j está na caixa i, temos $(n - 1)$ bolas para distribuir nas r caixas sem restrição quanto ao destino das bolas. Pode-se colocar a bola j na caixa i de uma maneira apenas, e as $(n - 1)$ bolas restantes podem ser colocadas nas r caixas de r^{n-1} maneiras. Assim a probabilidade desejada é $r^{n-1}/r^n = 1/r$.

Traduzindo em linguagem de amostragem aleatória, vemos que em uma amostra aleatória de tamanho n, extraída com reposição de uma população de r objetos, é igualmente provável que o j-ésimo elemento da amostra seja qualquer um dos r objetos.

As considerações acima estendem-se facilmente de uma caixa específica para k caixas, $1 \leqslant k \leqslant r$. Deixamos como exercício mostrar que a probabilidade de que k bolas específicas ocupem k caixas específicas é simplesmente r^{-k}. Em linguagem de amostragem aleatória este resultado diz que se extrairmos com reposição uma amostra de tamanho n de uma população de r objetos, a probabilidade de que o j_1-ésimo, j_2-ésimo, ... j_k-ésimo elementos da amostra sejam k objetos específicos quaisquer é r^{-k}.

Seja $A_j(i)$ o evento de que o j-ésimo elemento da amostra é o i-ésimo objeto. Então acabamos de dizer que para qualquer escolha $j_1 < j_2 < \cdots < j_k$, $1 \leqslant k \leqslant n$, de elementos de amostra (isto é, bolas) e qualquer escolha $i_1, i_2, \ldots, i_k$ de objetos (isto é, caixas), temos

$$P(A_{j_1}(i_1) \cap A_{j_2}(i_2) \cap \cdots \cap A_{j_k}(i_k)) = r^{-k}.$$

Como $P(A_j(i)) = r^{-1}$ para qualquer j e i, temos que

$$P(A_{j_1}(i_1) \cap \cdots \cap A_{j_k}(i_k)) = P(A_{j_1}(i_1)) \cdots P(A_{j_k}(i_k)). \tag{12}$$

Como esta relação é válida para todo k e todas as escolhas de $j_1, \ldots, j_k$, vemos que *os eventos* $A_1(i_1), \ldots, A_n(i_n)$ *são mutuamente independentes* para qualquer $i_1, i_2, \ldots, i_n$.

Se pensamos na extração aleatória de uma amostra de tamanho n de um conjunto de r objetos distintos como n repetições do experimento de escolher ao acaso um objeto deste mesmo conjunto, vemos que a afirmação de que os eventos $A_1(i_1), \ldots, A_n(i_n)$ são independentes significa que o resultado de um experimento não tem influência alguma sobre os resultados dos outros experimentos. Naturalmente isto está bem de acordo com a noção intuitiva de amostragem aleatória.

Exemplo 17. Suponha que se distribui aleatoriamente n bolas em r caixas. Obtenha a probabilidade de que tenha exatamente k bolas nas primeiras r_1 caixas.

Para resolver o problema observe que a probabilidade de que uma dada bola esteja em uma das primeiras r caixas, é r_1/r. Pense na distribuição de n bolas como n repetições do experimento de colocar uma bola em uma das r caixas. Suponha que o resultado é considerado sucesso se a bola cai em uma das primeiras r_1 caixas, e fracasso caso contrário. Então, do resultado da Seção 1.5., vemos que a probabilidade de que as primeiras r_1 caixas contenham exatamente k bolas é

$$\binom{n}{k}\left(\frac{r_1}{r}\right)^k\left(1-\frac{r_1}{r}\right)^{n-k}.$$

2.8. NÚMERO DE CAIXAS VAZIAS

Voltamos a considerar novamente o problema da distribuição aleatória de n bolas em r caixas e investigamos a probabilidade $p_k(r, n)$ de que exatamente k caixas estejam vazias.

Iniciamos a resolver o problema representando por A_i o evento de que a i-ésima caixa está vazia. Para que este evento ocorra, todas as n bolas devem estar nas $(r-1)$ caixas restantes, e isto pode acontecer de $(r-1)^n$ maneiras. Assim, $P(A_i) = (r-1)^n/r^n = (1-1/r)^n$.

De maneira semelhante, se $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant r$, o evento $A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}$ ocorre se, e somente se, todas as bolas estiverem nas $r-k$ caixas restantes. Conseqüentemente $P(A_{i_1} \cap \cdots \cap A_{i_k}) = (r-k)^n/r^n = (1-k/r)^n$. Podemos agora aplicar (7) para determinar a probabilidade de $A_1 \cup \cdots \cup A_n$ que é simplesmente o evento de que pelo menos uma caixa esteja vazia. Nesta situação $S_k = \binom{r}{k}(1-k/r)^n$, de modo que usando (7) obtemos

$$\begin{aligned} P(A_1 \cup \cdots \cup A_r) &= \sum_{k=1}^{r} (-1)^{k-1} S_k \\ &= \sum_{k=1}^{r} (-1)^{k-1} \binom{r}{k}\left(1-\frac{k}{r}\right)^n. \end{aligned}$$

Assim a probabilidade $p_0(r, n)$ de que todas as caixas estejam ocupadas é

$$\text{(13)} \qquad \begin{aligned} p_0(r,n) &= 1 - P(A_1 \cup \cdots \cup A_r) \\ &= 1 - \sum_{j=1}^{r} (-1)^{j-1}\binom{r}{j}\left(1-\frac{j}{r}\right)^n \\ &= \sum_{j=0}^{r} (-1)^{j}\binom{r}{j}\left(1-\frac{j}{r}\right)^n. \end{aligned}$$

Como etapa seguinte determinamos a probabilidade $\alpha_k(r, n)$ de que exatamente k caixas *específicas* (digamos as primeiras k caixas) estejam vazias. Este

evento pode ocorrer se todas as n bolas estiverem nas $(r - k)$ caixas restantes e se nenhuma dessas $r - k$ caixas estiverem vazias. O número de maneiras em que podemos distribuir n bolas em $(r - k)$ caixas de modo que nenhuma caixa fique vazia é $(r - k)^n p_0(r - k, n)$. Assim, a probabilidade desejada é

$$(14) \qquad \alpha_k(r, n) = \frac{(r - k)^n p_0(r - k, n)}{r^n} = \left(1 - \frac{k}{r}\right)^n p_0(r - k, n).$$

Podemos agora determinar facilmente as probabilidades $p_k(r, n)$. Para cada escolha de k números distintos $i_1, i_2, \ldots, i_k$ do conjunto de números $\{1, 2, \ldots, n\}$, o evento{exatamente k caixas $i_1, i_2, \ldots, i_k$ vazias}tem probabilidade $\alpha_k(r, n)$ e estes eventos são mutuamente disjuntos. Existem $\binom{r}{k}$ tais eventos e sua união é simplesmente o evento exatamente k caixas vazias. Assim

$$(15) \qquad p_k(r, n) = \binom{r}{k}\left(1 - \frac{k}{r}\right)^n p_0(r - k, n).$$

Usando a expressão de $p_0(r, n)$ dada em (13) vemos que

$$(16) \qquad p_k(r, n) = \binom{r}{k}\sum_{j=0}^{r-k}(-1)^j\binom{r-k}{j}\left(1 - \frac{j+k}{r}\right)^n.$$

Como acontece com o problema dos encontros, os problemas de ocupação têm diversas formulações. Apresentamos a seguir uma das mais famosas entre elas.

Problema do cupom. Coloca-se cupons ou, nos dias atuais, brinquedos nas caixas de cereais para atrair compradores infantis. Suponha que existem r tipos de cupons ou brinquedos e que é igualmente provável que uma dada caixa contenha qualquer um deles. Suponha que se adquire n caixas, determine a probabilidade de

(a) obter uma coleção com pelo menos um de cada tipo,

(b) não obter exatamente k dos r tipos de cupons ou brinquedos.

Exercícios

1. O código genético especifica um aminoácido através de uma seqüência de três nucleotídeos. Cada nucleotídeo pode ser de um dos quatro tipos T, A, C e G, sendo permitidas as repetições. Quantos aminoácidos podem ser codificados desta maneira?

2. O código Morse consiste de uma seqüência de pontos e traços em que repetições são permitidas.

   (a) Quantas letras se pode codificar usando exatamente n símbolos?

   (b) Qual é o número de letras que se pode codificar usando n ou menos símbolos?

3. Um homem possui n chaves das quais, exatamente uma abre a fechadura. Ele experimenta as chaves uma de cada vez, escolhendo ao acaso em cada tentativa uma das chaves que não foram experimentadas. Determine a probabilidade de que ele escolha a chave correta na r-ésima tentativa.

4. Um ônibus parte com 6 pessoas e para em 10 pontos diferentes. Supondo que os passageiros têm igual probabilidade de saltar em qualquer parada, determine a probabilidade de que dois passageiros não desembarquem na mesma parada.

5. Suponha que temos r caixas. Bolas são colocadas aleatoriamente nas caixas, uma de cada vez, até que alguma caixa contenha duas bolas pela primeira vez. Determine a probabilidade de que isto ocorra na n-ésima bola.

6. Uma caixa contém r bolas numeradas de 1 a r. Seleciona-se ao acaso N bolas da caixa (onde $N \leqslant r$), registra-se os seus números e repõe-se as N bolas na caixa. Se esse processo é repetido r vezes, qual é a probabilidade de que nenhuma das N bolas originais seja duplicada?

7. Se Sam e Peter estão entre n homens dispostos aleatoriamente em uma fila qual é a probabilidade de que haja exatamente k homens entre eles?

8. Um dominó é um bloco retangular dividido em dois sub-retângulos, como ilustra a figura abaixo. Cada sub-retângulo possui um número; sejam x e y esses números (não necessariamente distintos). Como o bloco é simétrico,

o dominó (x, y) é o mesmo que (y, x). Quantos blocos diferentes de dominó se pode fazer usando n números diferentes?

9. Considere o problema de encontros envolvendo n objetos e sejam i e r duas posições específicas distintas.

   (a) Qual é a probabilidade de que ocorra um encontro na posição i e não ocorra na posição r?

   (b) Dado que não ocorre um encontro na posição r, qual é a probabilidade de um encontro na posição i?

10. Suponha que se distribui n bolas em n caixas.

    (a) Qual é a probabilidade de que exatamente uma caixa esteja vazia? *Sugestão:* use o resultado do Exemplo 9.

    (b) Dado que a caixa 1 está vazia, qual a probabilidade de que somente uma caixa esteja vazia?

    (c) Dado que somente uma caixa está vazia, qual a probabilidade de que a caixa 1 esteja vazia?

11. Se distribuímos aleatoriamente n bolas em r caixas, qual é a probabilidade de que a caixa 1 contenha exatamente j bolas, $0 \leqslant j \leqslant n$?

12. Mostre que

$$\left(1 - \frac{n-1}{s}\right)^{n-1} \leq \frac{(s)_n}{s^n} \leq \left(1 - \frac{1}{s}\right)^{n-1}.$$

13. Uma caixa contém b bolas pretas e r bolas vermelhas. Bolas são extraídas sem reposição, uma de cada vez. Determine a probabilidade de obter a primeira bola preta na n-ésima extração.

O problema seguinte diz respeito a mãos de pôquer. Um baralho tem 52 cartas. Estas cartas consistem de 4 naipes chamadas paus, ouros, copas e espadas. Cada naipe tem 13 cartas com os símbolos 2, 3, 4, . . . , 10, J, Q, K, A. Uma mão de pôquer consiste de 5 cartas extraídas do baralho sem reposição e sem consideração de ordem. Considera-se que constituem seqüências as mãos dos seguintes tipos: A, 2, 3, 4, 5, 2, 3, 4, 5, 6; . . . ; 10, J, Q, K, A.

14. Determine a probabilidade de ocorrência de cada uma das seguintes mãos de pôquer:

    (a) Royal flush ((10, J, Q, K, A) do mesmo naipe);
    (b) Straight flush (cinco cartas do mesmo naipe em seqüência);
    (c) Four (valores da forma (x, x, x, x, y) onde x e y são distintos);
    (d) Full house (valores de forma (x, x, x, y, y) onde x e y são distintos);
    (e) Flush (cinco cartas do mesmo naipe);
    (f) Straight (cinco cartas em seqüência, sem consideração de naipes);
    (g) Trinca (valores de forma (x, x, x, y, z) onde x, y, e z são distintos);
    (h) Dois pares (valores da forma (x, x, y, y, z) onde x, y e z são distintos);
    (i) Um par (valores da forma (w, w, x, y, z) onde w, x, y e z são distintos).

15. Uma caixa contém dez bolas numeradas de 1 a 10. Seleciona-se uma amostra aleatória de 3 elementos. Determine a probabilidade de que as bolas 1 e 6 estejam entre as bolas selecionadas.

16. Extrai-se uma a uma as cartas de um baralho comum até obter o primeiro rei. Determine a probabilidade de que isto ocorra na n-ésima carta extraída.

17. Suponha que se extrai uma amostra de tamanho n de uma população de r elementos. Determine a probabilidade de que nenhum dos k elementos específicos esteja na amostra se o método utilizado é

    (a) amostragem sem reposição;
    (b) amostragem com reposição.

18. Suponha que se extrai, sem reposição, uma amostra de tamanho n de uma população de r elementos. Obtenha a probabilidade de que k objetos dados estejam incluídos na amostra.

19. Suponha que se permuta n objetos aleatoriamente entre si. Mostre que a probabilidade de que k objetos específicos ocupem k posições específicas é $(n-k)!/n!$.

20. Em relação ao Exemplo 14, mostre que

$$\frac{\binom{n}{k}\binom{r-n}{m-k}}{\binom{r}{m}} = \frac{\binom{m}{k}\binom{r-m}{n-k}}{\binom{r}{n}}$$

21. Uma caixa contém 40 fusíveis bons e 10 defeituosos. Supondo que se seleciona 10 fusíveis, qual é a probabilidade de que todos eles sejam bons?

22. Qual é a probabilidade que as mãos norte e sul de bridge juntas (um total de 26 cartas) contenham exatamente 3 ases?

23. Qual é a probabilidade de que de 4 cartas extraídas de um baralho, 2 sejam pretas e 2 vermelhas?

24. Determine a probabilidade de que uma mão de pôquer não contenha nenhuma carta inferior a 7, dado que contém pelo menos uma carta superior a 10, onde ases são consideradas cartas altas.

25. Se você possui 3 bilhetes de uma loteria para a qual se vendeu n bilhetes e existem 5 prêmios, qual a probabilidade de você ganhar pelo menos um prêmio?

26. Uma caixa de 100 retentores contém 5 itens defeituosos. Qual é a probabilidade de que dois retentores selecionados ao acaso (sem reposição) da caixa sejam ambos bons?

27. Considere duas caixas, cada uma com r bolas numeradas de 1 a r. Toma-se de cada caixa uma amostra aleatória sem reposição de tamanho $n \leqslant r$. Obtenha a probabilidade de que as amostras contenham exatamente k elementos em comum.

3

VARIÁVEIS ALEATÓRIAS DISCRETAS

Considere o experimento de lançar uma moeda três vezes em que a probabilidade de obter cara em um lançamento individual é p. Suponha que ganhemos \$ 1 dólar para cada lançamento que resulte em cara, mas percamos \$ 1 dólar para cada lançamento que resulte em coroa. Nesta situação, a quantidade de interesse é claramente o nosso ganho total. Seja X essa quantidade. É claro que X pode assumir somente os valores \$3, \$1, −\$1 e −\$3. Não podemos dizer com certeza qual destes será o valor de X, uma vez que este valor depende do resultado de nosso experimento aleatório. Se, por exemplo, o resultado for HHH, X será \$3; enquanto para o resultado HTH, X será \$1. Na tabela abaixo listamos os valores de X (em dólares) correspondentes a cada um dos oito possíveis resultados.

ω	$X(\omega)$	$P\{\omega\}$
HHH	3	p^3
HHT	1	$p^2(1-p)$
HTH	1	$p^2(1-p)$
THH	1	$p^2(1-p)$
HTT	−1	$p(1-p)^2$
THT	−1	$p(1-p)^2$
TTH	−1	$p(1-p)^2$
TTT	−3	$(1-p)^3$

Podemos pensar em X como uma função real no espaço de probabilidade correspondente ao experimento. Então, para cada $\omega \in \Omega$, $X(\omega)$ é um dos valores 3, 1, −1, −3. Considere, por exemplo, o evento $\{\omega: X(\omega) = 1\}$. Este conjunto contém os três pontos ω_2, ω_3 e ω_4 correspondentes aos resultados HHT, HTH e THH, respectivamente. A última coluna da tabela dá as probabilidades associadas

aos oito resultados possíveis do experimento. Desta tabela vemos que o evento $\{\omega: X(\omega) = 1\}$ tem probabilidade $3p^2(1-p)$. Geralmente expressamos este fato abreviadamente dizendo que $\{X = 1\}$ tem probabilidade $3p^2(1-p)$. Considerações similares naturalmente se aplicam a outros valores que X pode assumir. Vemos portanto que para cada valor possível de X existe uma probabilidade precisamente definida de X assumir esse valor. Como veremos na seção seguinte, a quantidade X é um exemplo do que se chama variável aleatória discreta.

3.1. DEFINIÇÕES

Seja $(\Omega, \mathscr{A}, P)$ um espaço arbitrário de probabilidade e seja X uma função real em Ω, tomando apenas um número finito ou infinito enumerável de valores $x_1, x_2, \ldots$ Como no exemplo que acabamos de dar, certamente gostaríamos de poder falar na probabilidade de que X assuma o valor x_i para cada i. Para isso, precisamos saber que $\{\omega \in \Omega: X(\omega) = x_i\}$ é um evento, isto é, um membro de $\mathscr{A}$ para cada i. Se, como no exemplo anterior, $\mathscr{A}$ é o σ-álgebra de todos os subconjuntos de Ω, este certamente é o caso. Pois neste caso, não importa qual seja x_i, $\{\omega: M(\omega) = x_i\}$ é um subconjunto de Ω e portanto um membro de $\mathscr{A}$, já que $\mathscr{A}$ contém todos os subconjuntos possíveis de Ω. Entretanto, como indicamos na Seção 1.2., em geral $\mathscr{A}$ não consiste de todos os subconjuntos de Ω, de modo que não temos nenhuma garantia de que $\{\omega \in \Omega: X(\omega) = x_i\}$ esteja em $\mathscr{A}$. A única saída razoável é *supor explicitamente* que X é uma função em Ω tal que satisfaça a condição desejada. Isto nos conduz à definição a seguir.

Definição 1. Uma variável aleatória discreta real X, em um espaço de probabilidade $(\Omega, \mathscr{A}, P)$, é uma função X cujo domínio é Ω e cujo contradomínio é um subconjunto finito ou infinito enumerável $\{x_1, x_2, \ldots\}$ dos números reais R tal que $\{\omega: X(\omega) = x_i\}$ é um evento para todo i.

Então $\{\omega: X(\omega) = x_i\}$ é por definição um evento e podemos falar de sua probabilidade. Por brevidade representamos em geral o evento $\{\omega: X(\omega) = x_i\}$ por $\{X = x_i\}$, e a probabilidade deste evento por $P(X = x_i)$ em vez de $P(\{\omega: X(\omega) = x_i\})$.

Seja X uma variável aleatória discreta real. Então $\{\omega: X(\omega) = x\}$ é um evento para qualquer número real x. Realmente, se $x_1, x_2, \ldots$ são os valores que X pode assumir, $\{\omega: X(\omega) = x_i\}$ é um evento de acordo com a definição de uma variável aleatória discreta real. Se x não é um desses valores, então $\{\omega: X(\omega) = x\} = \phi$ que também é um evento.

Se os valores possíveis de uma variável aleatória discreta X consistem apenas de números inteiros ou de números inteiros não-negativos, dizemos que X é uma variável aleatória inteira ou uma variável aleatória inteira não-negativa, respectivamente. A maioria das variáveis aleatórias discretas, que ocorrem nas aplicações, são inteiras não-negativas.

Definição 2. Chama-se função discreta de densidade X a uma função real f definida por $f(x) = P(X = x)$. Diz-se que um número real x é um valor possível de X se $f(x) > 0$.

Representaremos a função de densidade de X por f_X sempre que necessário para indicar que ela é função de densidade relativa à variável aleatória X.

Exemplo 1. Seja X a variável aleatória introduzida no começo deste capítulo na discussão de três lançamentos de uma moeda com, digamos, $p = 0{,}4$. Então X tem a densidade discreta f dada por

$$f(-3) = 0{,}216; \quad f(-1) = 0{,}432; \quad f(1) = 0{,}288; \quad f(3) = 0{,}064$$

e $f(x) = 0$ se $x \neq -3, -1, 1, 3$. Pode-se representar esta densidade por meio de um diagrama como ilustra a Figura 1

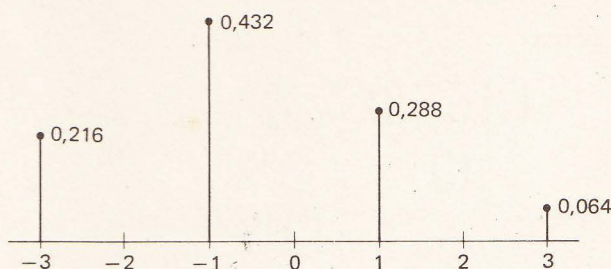



Figura 1

Exemplo 2. Distribuição binomial. Considere n repetições independentes de um experimento simples do tipo sucesso-fracasso discutido na Seção 1.5. Seja S_n o número de sucessos em n repetições. Então S_n é uma variável aleatória que pode assumir somente os valores $0, 1, 2, \ldots, n$. Mostramos no Capítulo 1 que, para um número inteiro k, $0 \leqslant k \leqslant n$,

$$P(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k};$$

portanto a densidade f de S_n é dada por

$$f(x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x}, & x = 0, 1, 2, \ldots, n, \\ 0, & \text{para outros valores de } x. \end{cases}$$

Esta densidade, que está entre as mais importantes que ocorrem na teoria da probabilidade, chama-se *densidade binomial* de parâmetros n e p. A densidade do Exemplo 1 é uma densidade binomial de parâmetros $n = 3$ e $p = 0{,}4$.

Refere-se freqüentemente a uma variável aleatória X, com densidade binomial, dizendo que X tem uma *distribuição* binomial (com parâmetros n e p, quando se deseja ser mais preciso). Usa-se também fraseologia semelhante para outras variáveis aleatórias que tem designação específica.

Como foi explicado no Capítulo 2, a distribuição binomial ocorre na amostragem aleatória com reposição. Na amostragem aleatória sem reposição, temos a distribuição discutida a seguir.

Exemplo 3. Distribuição hipergeométrica. Considere uma população de r objetos dos quais r_1, sã., de um tipo e $r_2 = r - r_1$ são de um segundó tipo. Suponha que se extraia desta população uma amostra aleatória sem reposição de tamanho $n \leqslant r$. Seja X o número de objetos do primeiro tipo na amostra. Então X é uma variável aleatória cujos valores possíveis são $0, 1, 2, \ldots, n$. Dos resultados da Seção 2.4. sabemos que

$$P(X = x) = \frac{\binom{r_1}{x}\binom{r - r_1}{n - x}}{\binom{r}{n}}, \qquad x = 0, 1, 2, \ldots, n.$$

Mas podemos escrever

$$\frac{\binom{r_1}{x}\binom{r - r_1}{n - x}}{\binom{r}{n}} = \frac{(r_1)_x(r - r_1)_{n-x}}{x!(n - x)!}\frac{n!}{(r)_n}$$

$$= \binom{n}{x}\frac{(r_1)_x(r - r_1)_{n-x}}{(r)_n}.$$

Assim podemos escrever a densidade f de X de duas formas

$$f(x) = \begin{cases} \dfrac{\binom{r_1}{x}\binom{r - r_1}{n - x}}{\binom{r}{n}}, & x = 0, 1, 2, \ldots, n, \\ 0, & \text{outros valores de } x \end{cases}$$

ou

$$f(x) = \begin{cases} \binom{n}{x}\dfrac{(r_1)_x(r - r_1)_{n-x}}{(r)_n}, & x = 0, 1, 2, \ldots, n, \\ 0, & \text{outros valores de } x \end{cases}$$

Esta densidade chama-se *densidade hipergeométrica*.

Apresentamos a seguir alguns exemplos mais de variáveis aleatórias.

Exemplo 4. Variável aleatória constante. Seja c um número real. Então a função X definida através de $X(\omega) = c$ para todo ω é uma variável aleatória discreta, já que $\{\omega\colon X(\omega) = c\}$ é todo o conjunto Ω e Ω é um evento. Claramente $P(X = c) = 1$, de modo que a densidade f de X é simplesmente $f(c) = 1$ e $f(x) = 0$, $x \neq c$. Esta variável aleatória chama-se variável constante. É sob este ponto de vista que uma constante numérica é considerada uma variável aleatória.

Exemplo 5. Variável aleatória indicadora. Seja A um evento. Faça $X(\omega) = 1$ se $\omega \in A$ e $X(\omega) = 0$ se $\omega \notin A$. Então o evento A ocorre se, e somente se,

$X = 1$. Esta variável aleatória chama-se variável aleatória indicadora de A porque o valor de X, nos diz se o evento A ocorre ou não.

Reciprocamente se X é uma variável aleatória em um espaço de probabilidade $(\Omega, \mathcal{A}, P)$, que toma o valor 1 ou 0, então X é a variável aleatória indicadora do evento.

$$A = \{ \omega : X(\omega) = 1 \}.$$

Seja $p = P(X = 1)$. A densidade f de X é dada por

$$f(0) = 1 - p, \quad f(1) = p \quad \text{e} \quad f(x) = 0 \quad \text{para} \quad x \neq 0 \quad \text{ou} \quad 1.$$

Exemplo 6. Considere o seguinte jogo de azar. Divide-se um alvo circular de raio 1 em n zonas por meio de n discos concêntricos de raios $1/n, 2/n, \ldots, n/n = 1$, como ilustra a Figura 2 para o caso $n = 5$. Lança-se aleatoriamente um dardo ao alvo, e se ele atinge a zona anular entre os círculos de raios i/n e $(i + 1)/n$, ganha-se i dólares, $i = 0, 1, 2, \ldots, n - 1$. Seja X a importância ganha. Obtenha a densidade de X.

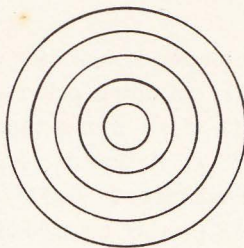



Figura 2

O espaço de probabilidade escolhido para este experimento será o espaço uniforme de probabilidade sobre o disco de raio 1. Claramente X é uma variável aleatória discreta neste espaço com os valores possíveis $1, 2, \ldots, n$. O evento $A = \{ X = n - i \}$ ocorre se, e somente se, o dardo atingir a região limitada pelos círculos de raios i/n e $(i + 1)/n$. De acordo com a discussão de Seção 1.2., a probabilidade de A é a área de A dividida pela área do disco unitário. Assim para $i = 0, 1, 2, \ldots, n - 1$,

$$P(X = n - i) = P(A)$$

$$= \frac{\pi \left[\left(\frac{i + 1}{n} \right)^2 - \left(\frac{i}{n} \right)^2 \right]}{\pi} = \frac{2i + 1}{n^2}.$$

Fazendo $n - i = x$, vemos que a densidade de X é

$$f(x) = \begin{cases} \dfrac{2(n - x) + 1}{n^2}, & x = 1, 2, \ldots, n, \\ 0, & \text{para outros valores de } x \end{cases}$$

A densidade f de uma variável aleatória discreta X tem as três seguintes importantes propriedades:

(i) $f(x) \geqslant 0, x \in R$.

(ii) $\{ x : f(x) \neq 0 \}$ *é um subconjunto finito ou infinito enumerável de* R. Seja $\{ x_1, x_2, \ldots \}$ *este conjunto. Então*

(iii) $\Sigma_i f(x_i) = 1$.

As propriedades (i) e (ii) são imediatas da definição de função de densidade discreta de X. Para verificar (iii), observe que os eventos $\{ \omega : X(\omega) = x_i \}$ são mutuamente disjuntos e sua união é Ω. Assim

$$\sum_i f(x_i) = \sum_i P(X = x_i)$$

$$= P\left(\bigcup_i \{X = x_i\}\right) = P(\Omega) = 1.$$

Definição 3. Diz-se que uma função real f definida em R é uma função de densidade discreta desde que ela satisfaça as propriedades (i), (ii) e (iii) mencionadas acima.

É fácil ver que qualquer função de densidade discreta é função de densidade de alguma variável aleatória X. Em outras palavras, dado f, podemos construir um espaço de probabilidade $(\Omega, \mathscr{A}, P)$ e uma variável aleatória X, definida em Ω, cuja densidade discreta é f. Realmente, suponha que f é dado e que $\{x_1, x_2, \ldots\}$ é o conjunto de valores onde $f(x) \neq 0$. Tome-se $\{ x_1, x_2, \ldots \}$ para Ω, todos os subconjuntos de Ω para $\mathscr{A}$ e a medida de probabilidade definida em $\mathscr{A}$ por $P(\omega) = f(x_i)$ se $\omega = x_i$ para P. Então a variável aleatória X definida por $X(\omega) = x_i$ se $\omega = x_i$ é uma variável aleatória com as características expostas acima. Isso pode ser constatado observado que $\{ \omega : X(\omega) = x_i \} = \{ x_i \}$ e assim

$$P(X = x_i) = P(\{x_i\}) = f(x_i).$$

O resultado acima nos garante que asserções como "seja X uma variável aleatória com densidade discreta f" sempre fazem sentido, mesmo que não especifiquemos diretamente o espaço de probabilidade sobre o qual X é definido. Para economia de escrita, usaremos doravante o termo densidade no lugar de densidade discreta até o final deste capítulo.

A noção de variável aleatória discreta constitui um meio conveniente de descrever um experimento aleatório que tem um número finito ou infinito enumerável de resultados possíveis. Não precisamos nos preocupar em estabelecer um espaço de probabilidade para o experimento. Em vez disso podemos simplesmente introduzir uma variável aleatória X que assuma os valores $x_1, x_2, \ldots$ tal que $X = x_i$ se, e somente se, o experimento conduz ao i-ésimo resultado. Assim, por exemplo, na extração aleatória de uma carta de um baralho com n cartas, podemos fazer $X = i$ quando a i-ésima carta é extraída. Então $P(X = i) = n^{-1}$, e poderíamos descrever o experimento dizendo que observamos uma variável aleatória X tomando os

valores inteiros 1, 2, . . ., n e tendo $f(x) = n^{-1}$ para $x = 1, 2, \ldots, n$ e $f(x) = 0$ para outros valores de x como sua função de densidade.

Pode-se em geral descrever a realização de um experimento que tem um número finito ou infinito enumerável de resultados possíveis como a observação do valor de uma variável aleatória discreta. Na realidade, muitas vezes esta é a forma como o experimento nos apresenta, e freqüentemente é mais fácil pensar no experimento nestes termos do que em termos de um espaço de probabilidade.

Como ilustração dessa idéia, considere o experimento de selecionar ao acaso um ponto do subconjunto S de R, consistindo dos pontos distintos $x_1, x_2, \ldots, x_s$. Então a função f definida por

$$f(x) = \begin{cases} s^{-1}, & x = x_1, x_2, \ldots, x_s, \\ 0, & \text{para outros valores de } x. \end{cases}$$

é claramente uma função de densidade discreta. Quando uma variável aleatória X tem esta densidade, dizemos que ela se *distribui uniformemente em* S. Observar um valor de X correspondente à noção intuitiva de escolher ao acaso um ponto de S.

Apresentaremos agora mais duas densidades discretas que serão muito úteis na resolução de certas classes de problemas cuja importância se tornará aparente mais tarde.

Exemplo 7. Densidades geométricas. Seja $0 < p < 1$. Então a função real f definida em R por

$$f(x) = \begin{cases} p(1-p)^x, & x = 0, 1, 2, \ldots, \\ 0, & \text{para outros valores de } x. \end{cases}$$

é uma função de densidade discreta chamada *densidade geométrica* de parâmetro p.

Para ver que f é uma densidade, tudo que precisamos fazer é verificar se ela satisfaz a condição (iii), pois obviamente satisfaz as condições (i) e (ii). Mas (iii) decorre do conhecido fato de que a soma da série geométrica $\sum_{x=0}^{\infty} (1-p)^x$ é simplesmente p^{-1}.

Exemplo 8. Densidades binomiais negativas. Seja α um número real positivo qualquer e seja $0 < p < 1$. Uma densidade intimamente relacionada com a geométrica é a *densidade binomial negativa* de parâmetros α e p definida por

$$(1) \qquad f(x) = \begin{cases} p^\alpha \binom{-\alpha}{x} (-1)^x (1-p)^x, & x = 0, 1, 2, \ldots, \\ 0, & \text{para outros valores de } x. \end{cases}$$

Para mostrar que isto é uma densidade, devemos mostrar que (1) satisfaz as propriedades (i)–(iii). Obviamente a propriedade (ii) é satisfeita. Para ver que (1) satisfaz (i), observamos que para um número inteiro não-negativo x temos

$$\binom{-\alpha}{x} = \frac{(-\alpha)_x}{x!} = \frac{(-\alpha)(-\alpha-1)\cdots(-\alpha-x+1)}{x!}$$

$$= \frac{(-1)^x(\alpha)(\alpha+1)\cdots(\alpha+x-1)}{x!}$$

$$= (-1)^x \frac{(\alpha+x-1)_x}{x!}$$

$$= (-1)^x \binom{\alpha+x-1}{x}.$$

Assim

$$(2) \qquad p^\alpha \binom{-\alpha}{x}(-1)^x(1-p)^x = p^\alpha \binom{\alpha+x-1}{x}(1-p)^x.$$

Como o segundo membro de (2) é claramente não-negativo, vemos que (i) é verdadeiro.

Para verificar (iii), lembre-se que a série de Taylor de $(1-t)^{-\alpha}$ para $-1 < t < 1$ é

$$(3) \qquad (1-t)^{-\alpha} = \sum_{x=0}^{\infty} \binom{-\alpha}{x}(-t)^x.$$

De (3) com $t = 1 - p$ vemos que

$$p^{-\alpha} = \sum_{x=0}^{\infty} \binom{-\alpha}{x}(-1)^x(1-p)^x$$

e portanto que $\Sigma_x f(x) = 1$.

De (2) vemos que podemos escrever a densidade binomial negativa na forma alternativa

$$(4) \qquad f(x) = \begin{cases} p^\alpha \binom{\alpha+x-1}{x}(1-p)^x, & x = 0, 1, 2, \ldots, \\ 0, & \text{para outros valores de } x. \end{cases}$$

Para alguns propósitos esta forma é mais útil do que a dada em (1). Observe que a densidade geométrica de parâmetro p é a densidade binomial negativa de parâmetros $\alpha = 1$ e p.

Exemplo 9. Densidades de Poisson. Seja λ um número positivo. Define-se a densidade de Poisson de parâmetro λ através de

$$f(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, \ldots, \\ 0, & \text{para outros valores de } x. \end{cases}$$

É evidente que esta função satisfaz as propriedades (i) e (ii) de definição de função de densidade discreta. A propriedade (iii) segue-se imediatamente da expansão em série de Taylor da função exponencial, a saber,

$$e^\lambda = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}.$$

Sabe-se de experiência que muitos fenômenos aleatórios que envolvem contagem seguem aproximadamente a distribuição de Poisson. Alguns exemplos de

tais fenômenos são o número de átomos de uma substância radioativa que se desintegram na unidade de tempo, o número de chamadas que chegam a uma central telefônica na unidade de tempo, o número de erros de impressão por página de um livro e o número de colônias de bactérias em um recipiente de petri untado com uma suspensão de bactérias. Um tratamento completo desses modelos requer a noção de processo de Poisson, que será discutido no Capítulo 9.

3.2. CÁLCULOS COM DENSIDADES

Até aqui restringimos nossa atenção à determinação de $P(X = x)$. Freqüentemente estamos interessados em determinar a probabilidade de $\{\omega: X(\omega) \in A\}$ onde A é um subconjunto de R com mais de um ponto.

Seja A um subconjunto de R e seja X uma variável aleatória discreta tendo valores possíveis distintos $x_1, x_2, \ldots$ Então $\{\omega: X(\omega) \in A\}$ é um evento. Para ver isso, observe que

$$\{\omega \mid X(\omega) \in A\} = \bigcup_{x_i \in A} \{\omega \mid X(\omega) = x_i\}, \tag{5}$$

onde $\bigcup_{x_i \in A}$ representa a união sobre todo valor de i, tal que $x_i \in A$. Geralmente abrevia-se o evento $\{\omega: X(\omega) \in A\}$ para $\{X \in A\}$ e representa-se sua probabilidade por $P(X \in A)$. Se $-\infty \leqslant a < b \leqslant \infty$ e A é um intervalo com extremos em a e b, digamos $A = (a, b]$, então geralmente escrevemos $P(a < X \leqslant b)$ em vez de $P(X \in (a, b])$. Usa-se notação semelhante para outros intervalos com estes pontos extremos.

Usa-se também uma notação abreviada para probabilidades condicionais. Assim, por exemplo, se A e B são dois subconjuntos de R, escrevemos $P(X \in A \mid X \in B)$ para a probabilidade condicional do evento $\{X \in A\}$ dado o evento $\{X \in B\}$.

Seja f a densidade de X. Podemos determinar $P(X \in A)$ diretamente da densidade f por meio da fórmula

$$P(X \in A) = \sum_{x_i \in A} f(x_i), \tag{6}$$

onde $\Sigma_{x_i \in A}$ representa a soma sobre todo i tal que $x_i \in A$. Esta fórmula segue-se imediatamente de (5) já que os eventos $\{\omega \mid X(\omega) = x_i\}$, $i = 1, 2, \ldots,$ são disjuntos. Geralmente abrevia-se o segundo membro de (6) para $\Sigma_{x \in A} f(x)$. Em termos desta notação, (6) transforma-se em

$$P(X \in A) = \sum_{x \in A} f(x). \tag{7}$$

A função $F(t)$, $-\infty < t < \infty$, definida por

$$F(t) = P(X \leq t) = \sum_{x \leq t} f(x), \qquad -\infty < t < \infty,$$

chama-se *função de distribuição* da variável aleatória X ou da densidade f. Segue-se imediatamente da definição de função de distribuição que

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a).$$

Se X é uma variável aleatória inteira, então

$$F(t) = \sum_{x=-\infty}^{[t]} f(x),$$

onde $[t]$ representa o maior inteiro menor ou igual a t (por exemplo, $[2, 6] = [2] = 2$). Vemos que F é uma função não-decrescente e que, para qualquer inteiro x, F dá um salto de magnitude $f(x)$ no ponto x e F é constante no intervalo $[x, x + 1)$. Propriedades adicionais das funções de distribuição serão obtidas de um ponto de vista mais geral no Capítulo 5.

Exemplo 10. Seja $S = \{1, 2, \ldots, 10\}$ e seja X uma variável aleatória uniformemente distribuída em S. Então $f(x) = 1/10$ para $x = 1, 2, \ldots, 10$ e $f(x) = 0$ para outros valores de x. A função de distribuição de X é dada por $F(t) = 0$ para $t < 1$, $F(t) = 1$ para $t > 10$ e

$$F(t) = \sum_{x=1}^{[t]} f(x) = \frac{[t]}{10}, \qquad 1 \leq x \leq 10.$$

O gráfico da função $F(t)$ é dado na Figura 3. A probabilidade $P(3 < X \leqslant 5)$ pode ser calculada como

$$P(3 < X \leq 5) = f(4) + f(5) = 2/10$$

ou como

$$P(3 < X \leq 5) = F(5) - F(3) = 5/10 - 3/10 = 2/10.$$

Analogamente $P(3 \leqslant X \leqslant 5)$ obtém-se como

$$P(3 \leq X \leq 5) = f(3) + f(4) + f(5) = 3/10$$

ou como

$$P(3 \leq X \leq 5) = P(2 < X \leq 5) = F(5) - F(2) = 5/10 - 2/10 = 3/10.$$

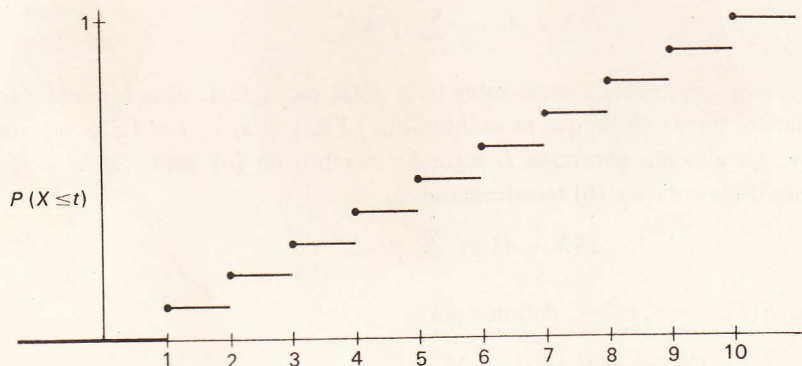



Figura 3

Exemplo 11. Seja X uma variável aleatória com distribuição geométrica de parâmetro p. Obtenha a função de distribuição de X e determine $P(X \geqslant x)$ para um número inteiro não-negativo de x.

A densidade de X, de acordo com o Exemplo 7, é

$$f(x) = \begin{cases} p(1-p)^x, & x = 0, 1, 2, \ldots, \\ 0, & \text{outros valores de } x. \end{cases}$$

Assim $F(t) = 0$ para $t < 0$ e

$$F(t) = \sum_{x=0}^{[t]} p(1-p)^x, \qquad t \geq 0.$$

Usando a fórmula para a soma de uma progressão geométrica finita, concluímos que

$$F(t) = 1 - (1-p)^{[t]+1}, \qquad t \geq 0.$$

Em particular, para um inteiro não-negativo x, $F(x-1) = 1 - (1-p)^x$ e, portanto,

$$\begin{aligned} P(X \geq x) &= 1 - P(X < x) = 1 - P(X \leq x - 1) \\ &= 1 - F(x-1) = (1-p)^x. \end{aligned}$$

Variáveis aleatórias com distribuição geométrica ocorrem naturalmente nas aplicações. Suponha que temos um componente de um equipamento, tal como um fusível elétrico, que não deteriora nem melhora com o passar do tempo, mas pode falhar devido a causas esporádicas e aleatórias que ocorrem homogeneamente no tempo. Suponha que se observe o objeto a intervalos fixos de tempo, tais como horas ou dias, e seja X o número de unidades de tempo até e inclusive a primeira falha, suponha que o objeto está novo no tempo 0. Claramente X é uma variável aleatória discreta cujos valores possíveis são os inteiros $1, 2, \ldots$ O evento $\{X = n\}$ ocorre se, e somente se, o objeto falha pela primeira vez no n-ésimo período de tempo. Pode-se formular precisamente a noção de que o objeto não deteriora nem melhora com o tempo da seguinte forma. Se sabemos que o objeto não falhou até o tempo n de modo que $X > n$, a probabilidade de que o objeto não falhe até o tempo $n + m$, isto é, $P(X > n + m/X > n)$, deve ser igual a probabilidade de um objeto novo instalado no tempo n, não falhar até o tempo $n + m$. Pode-se tomar o fato de que as causas da falha ocorrem homogeneamente no tempo como significando que esta probabilidade depende apenas do número de intervalos que decorrem entre n e $n + m$, isto é, m, mas não de n. Assim $P(X > n)$ deve satisfazer a equação

$$P(X > n + m \mid X > n) = P(X > m). \tag{8}$$

Uma vez que

$$P(X > n + m \mid X > n) = \frac{P(X > n + m)}{P(X > n)},$$

podemos reescrever (8) como

$$P(X > n + m) = P(X > n)P(X > m), \qquad n, m = 0, 1, 2, \ldots. \tag{9}$$

Fazendo $n = m = 0$ vemos que $P(X > 0 = P(X > 0)^2$, de modo que $P(X > 0)$ é igual a 1 ou 0. Se $P(X > 0) = 0$, então $P(X = 0) = 1$, o que é impossível no nosso caso já que X pode assumir somente valores inteiros positivos. Portanto $P(X > 0) = 1$.

Fazendo $p = P(X = 1)$, de modo que $P(X > 1) = 1 - p$, vemos de (9) que

$$P(X > n + 1) = (1 - p)P(X > n).$$

Por iteração sobre n segue-se que $P(X > n) = (1 - p)^n$. Assim para $n = 1, 2, \ldots,$

$$(10) \qquad P(X = n) = P(X > n - 1) - P(X > n)$$
$$= (1 - p)^{n-1} - (1 - p)^n = p(1 - p)^{n-1}.$$

Se $p = 0$, então $P(X = n) = 0$ para $n = 1, 2, \ldots$ e assim $P(X = +\infty) = 1$, isto é, o objeto nunca falha. Nós excluímos este caso de consideração. Da mesma forma excluímos $p = 1$ porque neste caso $P(X = 1) = 1$, e o objeto sempre falha no primeiro período de funcionamento.

Seja $Y = X - 1$. Então Y assume os valores $0, 1, 2, \ldots$ com probabilidades $P(Y = n) = p(1 - p)^n$. Vemos portanto que Y tem distribuição geométrica de parâmetro p.

Como acabamos de mostrar, a variável aleatória $Y = X - 1$ distribui-se geometricamente. Este exemplo é típico no sentido de que variáveis aleatórias com distribuição geométrica ocorrem geralmente em relação ao tempo de espera até a ocorrência de algum evento. Este fato será discutido mais detalhadamente depois de tratarmos das repetições independentes de um experimento na Seção 3.4.

3.3. VETORES ALEATÓRIOS DISCRETOS

Acontece, freqüentemente, estarmos interessados em estudar a relação entre duas ou mais variáveis aleatórias. Assim, por exemplo, na extração de uma amostra aleatória de tamanho n de uma caixa, contendo r bolas, numeradas de 1 a r, poderíamos desejar saber qual o número mais alto Y ou o número mais baixo Z observado entre as bolas selecionadas.

Seja $(\Omega, \mathscr{A}, P)$ um espaço de probabilidade e sejam $X_1, X_2, \ldots, X_r$ r variáveis aleatórias discretas definidas neste espaço. Então para cada ponto $\omega \in \Omega$ cada uma das variáveis aleatórias $X_1, \ldots, X_r$ assume um de seus valores possíveis, fato que será indicado escrevendo

$$X_1(\omega) = x_1, X_2(\omega) = x_2, \ldots, X_r(\omega) = x_r.$$

Em vez de pensar em observar r números reais $x_1, x_2, \ldots, x_r$, podemos pensar em observar *uma* r-tupla $\mathbf{x} = (x_1, x_2, \ldots, x_r)$, onde para cada i, x_i é um dentre o número finito ou infinito enumerável de valores que a variável aleatória X_i pode assumir.

Seja R^r a coleção de todas as r-tuplas de números reais. Um ponto $\mathbf{x} = (x_1, \ldots, x_r)$ de R^r é habitualmente chamado de *vetor r-dimensional*. Assim, para cada $\omega \in \Omega$, os r valores $X_1(\omega), \ldots, X_r(\omega)$ definem um ponto

$$\mathbf{X}(\omega) = (X_1(\omega), X_2(\omega), \ldots, X_r(\omega))$$

de R^r. Isto define uma função vetorial r-dimensional em Ω, $\mathbf{X}: \Omega \to R^r$, que se representa habitualmente por $\mathbf{X} = (X_1, X_2, \ldots, X_r)$. A função $\mathbf{X}$ é chamada de *vetor aleatório discreto r-dimensional*.

Acabamos de definir um vetor aleatório r-dimensional em termos de r variáveis aleatórias reais. Alternativamente, pode-se definir um vetor aleatório r-dimensional diretamente com a função $\mathbf{X}: \Omega \to R^r$, estendendo quase literalmente a definição de uma variável aleatória real.

Definição 4. Um vetor aleatório r-dimensional $\mathbf{X}$, é uma função $\mathbf{X}$ de Ω para R^r, assumindo um número finito ou infinito e numerável de valores $\mathbf{x}_1, \mathbf{x}_2, \ldots$ tais que

$$\{\omega: \mathbf{X}(\omega) = \mathbf{x}_i\}$$

é um evento para todo i.

Define-se a função de densidade f de um vetor aleatório $\mathbf{X}$ através de

$$f(x_1, \ldots, x_r) = P(X_1 = x_1, \ldots, X_r = x_r)$$

ou equivalentemente

$$f(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}), \qquad \mathbf{x} \in R^r.$$

Pode-se determinar a probabilidade de que $\mathbf{X}$ pertença ao subconjunto A de R^r usando o análogo de (7), a saber,

$$P(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A} f(\mathbf{x}).$$

Como no caso unidimensional, a função f tem as seguintes propriedades:

(i) $f(\mathbf{x}) \geqslant 0, \mathbf{x} \in R^r$.
(ii) $\{\mathbf{x}: f(\mathbf{x}) \neq 0\}$ *é um subconjunto finito ou infinito enumerável de* R^r, *que será representado por* $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$.
(iii) $\Sigma_i f(\mathbf{x}_i) = 1$.

Qualquer função real f definida em R^r, que tem estas três propriedades, será chamada de função de densidade discreta r-dimensional. O argumento apresentado para o caso unidimensional aplica-se literalmente para mostrar que qualquer função de densidade discreta r-dimensional é função de densidade de algum vetor aleatório r-dimensional.

Existe uma terminologia tradicional associada com vetores aleatórios e suas funções de densidade. Seja $\mathbf{X} = (X_1, X_2, \ldots, X_r)$ um vetor aleatório r-dimensional

com densidade f. Então a função f é habitualmente chamada de *função de densidade* conjunta das variáveis aleatórias $X_1, X_2, \ldots, X_r$. A função de densidade da variável aleatória X_i é então chamada de i-ésima *densidade marginal* de **X** ou de f.

Sejam X e Y duas variáveis aleatórias discretas. Para quaisquer números reais x e y, o conjunto $\{ \omega \mid X(\omega) = x \text{ e } Y(\omega) = y \}$ é um evento, que será representado por $\{ X = x, Y = y \}$. Suponha que os valores possíveis distintos de X sejam $x_1, x_2, \ldots$ e os de Y sejam $y_1, y_2, \ldots$ Para cada x, os eventos $\{X = x, Y = y_j\}$, $j = 1, 2, \ldots$, são disjuntos e sua união é o evento $\{ X = x \}$. Assim

$$P(X = x) = P\left(\bigcup_j \{X = x, Y = y_j\}\right)$$
$$= \sum_j P(X = x, Y = y_j) = \sum_y P(X = x, Y = y).$$

Esta última expressão resulta do uso da mesma convenção notacional introduzida para variáveis aleatórias na Seção 3.2.

Analogamente,

$$P(Y = y) = P\left(\bigcup_i \{X = x_i, Y = y\}\right)$$
$$= \sum_i P(X = x_i, Y = y) = \sum_x P(X = x, Y = y).$$

Em outras palavras, se conhecemos a densidade conjunta de duas variáveis aleatórias discretas X e Y, podemos obter a densidade f_X de X somando sobre y e, a densidade f_Y de Y somando sobre x. Assim em termos das densidades, se f é a densidade conjunta de X e Y, então

$$(11) \qquad f_X(x) = \sum_y f(x, y)$$

e

$$(12) \qquad f_Y(y) = \sum_x f(x, y).$$

Exemplo 12. Suponha que se extraia aleatoriamente, sem reposição, duas cartas de um baralho de 3 cartas numeradas de 1 a 3. Seja X o número da primeira carta e Y o da segunda. Então a densidade conjunta de X e Y é dada por $f(1, 2) = f(1, 3) = f(2, 1) = f(2, 3) = f(3, 1) = f(3, 2) = 1/6$ e $f(x, y) = 0$ para outros pares de x e y. A primeira densidade marginal, isto é, a densidade de X é dada por $f_X(1) = f(1, 1) + f(1, 2) + f(1, 3) = 0 + 1/6 + 1/6 = 2/6 = 1/3$ e similarmente para $x = 2$ e 3. Portanto $f_X(x) = 1/3, x = 1, 2, 3,$ e $f_X(x) = 0$ para outros valores de x, como era de se esperar.

Exemplo 13. Suponha que X e Y são duas variáveis aleatórias que assumem os valores de x e y, onde $x = 1, 2$ e $y = 1, 2, 3, 4$, com as probabilidades dadas na tabela abaixo.

$x \backslash y$	1	2	3	4
1	1/4	1/8	1/16	1/16
2	1/16	1/16	1/4	1/8

Então $f_X(1) = \Sigma_{y=1}^{4} f(1, y) = 1/4 + 1/8 + 1/16 + 1/16 = 1/2$, e $f_X(2) = 1 - f_X(1) = 1/2$ de modo que X tem distribuição uniforme sobre 1 e 2. Analogamente

$$f_Y(1) = 1/4 + 1/16 = 5/16,\ f_Y(2) = 3/16, f_Y(3) = 5/16, f_Y(4) = 3/16.$$

3.4. VARIÁVEIS ALEATÓRIAS INDEPENDENTES

Considere o experimento de lançar uma moeda e um dado. Acreditamos intuitivamente que o resultado do lançamento da moeda, qualquer que seja ele, não deve ter influência sobre o lançamento do dado, e vice-versa. Desejamos construir um modelo de probabilidade que reflita essas idéias. Seja X uma variável aleatória que é 1 ou 0, dependendo de se o lançamento da moeda resulta em cara ou coroa, isto é, o evento $\{X = 1\}$ representa o resultado de se observar cara e o evento $\{X = 0\}$ o de se observar coroa. De forma semelhante representamos o resultado do lançamento do dado por uma variável aleatória Y que assume o valor 1, 2, . . . ou 6 dependendo de, se o número observado no dado é 1, 2, . . . ou 6. Pode-se representar o resultado do experimento composto através do vetor aleatório (X, Y). A noção intuitiva de que os resultados observados, na moeda e no dado, não têm influência um sobre o outro pode ser traduzida, precisamente dizendo, que se x é um dos números 1 e 0 e y é um dos números 1, 2, . . . , 6, então os eventos $\{X = x\}$ e $\{X = y\}$ devem ser independentes. Assim o vetor aleatório (X, Y) deve ter a densidade conjunta $f(x, y)$ dada por

$$f(x,y) = \begin{cases} P(X = x)P(Y = y), & x = 0,1,\ y = 1, 2, \ldots, 6, \\ 0, & \text{outros valores de } x \text{ e } y. \end{cases}$$

Em outras palavras, a densidade conjunta f de X e Y deve ser dada por

$$f(x, y) = f_X(x) f_Y(y).$$

Definição 5. Sejam $X_1, X_2, \ldots, X_r$ r variáveis aleatórias discretas tendo as densidades $f_1, f_2, \ldots, f_r$, respectivamente. Diz-se que estas variáveis aleatórias são mutuamente independentes se a sua função de densidade conjunta f é dada por

(13) $$f(x_1, x_2, \ldots, x_r) = f_1(x_1) f_2(x_2) \cdots f_r(x_r).$$

Diz-se que as variáveis aleatórias são dependentes se elas não forem independentes. Como no caso do experimento composto de lançar uma moeda e um dado, a noção de variáveis aleatórias independentes constitui uma forma conveniente de formular precisamente a noção intuitiva de que os experimentos são independentes uns dos outros.

Considere duas variáveis aleatórias independentes com densidades f_X e f_Y, respectivamente. Então para dois subconjuntos quaisquer A e B de R.

(14) $$P(X \in A, Y \in B) = P(X \in A) P(Y \in B).$$

Para verificar (14), observe que

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{x \in A} \sum_{y \in B} f_{X,Y}(x, y) \\ &= \sum_{x \in A} \sum_{y \in B} f_X(x) f_Y(y) \\ &= \left[\sum_{x \in A} f_X(x) \right] \left[\sum_{y \in B} f_Y(y) \right] \\ &= P(X \in A) P(Y \in B). \end{aligned}$$

A fórmula (14) acima estende-se facilmente de 2 para r variáveis aleatórias independentes. Assim, se $A_1, A_2, \ldots A_r$ são r subconjuntos quaisquer de R, então

(15) $$P(X_1 \in A_1, \ldots, X_r \in A_r) = P(X_1 \in A_1) \cdots P(X_r \in A_r).$$

Exemplo 14. Sejam X e Y duas variáveis aleatórias independentes com distribuição geométrica de parâmetro p.

(a) Determine a distribuição de $\min(X, Y)$.
(b) Determine $P(\min(X, Y) = X) = P(Y \geqslant X)$.
(c) Determine a distribuição de $X + Y$.
(d) Determine $P(Y = y \mid X + Y = z)$ para $y = 0, 1, \ldots, z$.

Para resolver (a), observamos que para um inteiro não-negativo z

$$P(\min(X, Y) \geqslant z) = P(X \geqslant z, Y \geqslant z) = P(X \geqslant z) P(Y \geqslant z),$$

de modo que, pelo Exemplo 11

$$P(\min(X, Y) \geqslant z) = (1 - p)^z (1 - p)^z = (1 - p)^{2z}.$$

Segue-se do Exemplo 11 que $\min(X, Y)$ tem uma distribuição geométrica de parâmetro

$$1 - (1 - p)^2 = 2p - p^2.$$

Para resolver (b), observamos que

$$P(Y \geq X) = \sum_{x=0}^{\infty} P(X = x, Y \geq X)$$

$$= \sum_{x=0}^{\infty} P(X = x, Y \geq x)$$

$$= \sum_{x=0}^{\infty} P(X = x)P(Y \geq x)$$

$$= \sum_{x=0}^{\infty} p(1 - p)^x(1 - p)^x$$

$$= p \sum_{x=0}^{\infty} (1 - p)^{2x}$$

$$= p/(1 - (1 - p)^2) = p/(2p - p^2).$$

Para resolver (c), seja z um inteiro não-negativo. Então

$$P(X + Y = z) = \sum_{x=0}^{z} P(X = x, X + Y = z)$$

$$= \sum_{x=0}^{z} P(X = x, Y = z - x)$$

$$= \sum_{x=0}^{z} P(X = x)P(Y = z - x)$$

$$= \sum_{x=0}^{z} p(1 - p)^x p(1 - p)^{z-x}$$

$$= (z + 1)p^2(1 - p)^z.$$

A solução de (d), é dada por

$$P(Y = y \mid X + Y = z) = \frac{P(Y = y, X + Y = z)}{P(X + Y = z)}$$

$$= \frac{P(X = z - y, Y = y)}{P(X + Y = z)}$$

$$= \frac{P(X = z - y)P(Y = y)}{P(X + Y = z)}$$

$$= \frac{p(1 - p)^{z-y}p(1 - p)^y}{(z + 1)p^2(1 - p)^z}$$

$$= \frac{1}{z + 1}.$$

Considere um experimento (tal como o lançamento de um dado) que tenha somente um número finito ou infinito enumerável de resultados possíveis. Então, como já foi explicado, podemos pensar neste experimento como a observação do valor de uma variável aleatória X. Suponha que o experimento seja repetido n vezes. Pode-se descrever o experimento composto como a observação dos valores das variáveis aleatórias $X_1, X_2, \ldots, X_n$, onde X_i é o resultado do i-ésimo experimento. Repetindo-se os experimentos sob condições idênticas, presumivelmente o mecanismo de chance permanece o mesmo, de modo que devemos exigir que estas n variáveis aleatórias tenham a mesma densidade. Pode-se agora formular a noção intuitiva de que os repetidos experimentos não têm influência uns sobre os outros, exigindo que as variáveis aleatórias $X_1, X_2, \ldots, X_n$ sejam mutuamente independentes. Assim, em resumo, pode-se usar n variáveis aleatórias independentes, com densidade discreta f para representar n repetições de um experimento, tendo um número finito ou infinito enumerável de resultados possíveis.

Os experimentos aleatórios mais simples são aqueles que têm apenas dois resultados possíveis, que podemos designar de sucesso e fracasso. No lançamento de uma moeda, por exemplo, podemos pensar na obtenção de cara como sucesso, enquanto na extração de uma carta de um baralho podemos considerar a obtenção de um ás como sucesso. Suponha que realizamos n repetições do nosso experimento simples. Então podemos descrever a situação supondo que $X_1, X_2, \ldots, X_n$ são n variáveis aleatórias indicadoras tais que $X_i = 1$ ou 0 dependendo de se a i-ésima repetição do experimento resulta em sucesso ou fracasso. Na literatura, provas que podem resultar em sucesso ou fracasso são chamadas de *provas de Bernoulli*, e descreve-se a situação acima dizendo que realizamos n provas de Bernoulli com probabilidade comum de sucesso $p = P(X_i = 1)$. Neste contexto diz-se que uma variável aleatória que assume os valores 1 e 0, com probabilidades p e $1 - p$, tem uma *densidade de Bernoulli* de parâmetro p.

O resultado da realização de n provas de Bernoulli pode ser representado pelo vetor aleatório $\mathbf{X} = (X_1, X_2, \ldots, X_n)$. A informação contida neste vetor diz exatamente quais são as provas que resultaram em sucesso e quais resultaram em fracasso. Freqüentemente não se requer uma informação tão precisa, e tudo que desejamos conhecer é o número de sucessos S_n obtidos em n provas. Mostramos no Exemplo 2 que S_n se distribui binomialmente com parâmetros n e p. Observe que $S_n = X_1 + X_2 + \cdots + X_n$. Pode-se pensar em qualquer variável aleatória que se distribui binomialmente com esses parâmetros como sendo a soma de n variáveis aleatórias independentes de Bernoulli, cada uma com parâmetro p.

Consideremos agora as repetições independentes de um experimento que tem um número finito $r \geqslant 2$ de resultados possíveis.

3.4.1. A DISTRIBUIÇÃO MULTINOMIAL. Considere um experimento, tal como o lançamento de um dado, que pode resultar em somente um número finito r de distintos resultados possíveis. Podemos representar este experimento dizendo que

observamos uma variável aleatória Y que assume os valores $1, 2, \ldots, r$, de modo que o evento $\{ Y = i \}$ representa o fato de que o experimento produziu o i-ésimo resultado. Seja $p_i = P(Y = i)$. Se realizamos n repetições independentes do experimento, podemos representar o resultado dessas n provas por meio de um vetor aleatório n-dimensional $(Y_1, \ldots, Y_n)$, onde a variável aleatória Y_j corresponde a j-ésima prova. Aqui as variáveis aleatórias $Y_1, \ldots, Y_n$ são mutuamente independentes e $P(Y_j = i) = p_i$.

O vetor aleatório $(Y_1, \ldots, Y_n)$ nos dá o resultado de cada uma dessas n provas. Como no caso de $r = 2$ resultados, freqüentemente não estamos interessados numa descrição tão detalhada, mas gostaríamos de saber apenas quantas das n provas conduziram a cada um dos diversos resultados possíveis. Seja X_i, $i = 1, 2, \ldots, r$, o número de provas que produzem o i-ésimo resultado. Então $X_i = x_i$ se, e somente se, exatamente x_i das n variáveis aleatórias $Y_1, \ldots, Y_n$ assumem o valor i, isto é, exatamente x_i das n provas produzem o i-ésimo resultado.

Por exemplo, para $r = 3$ e $n = 5$, se

$$Y_1 = 2, Y_2 = 3, Y_3 = 3, Y_4 = 2, \text{ e } Y_5 = 2,$$

então

$$X_1 = 0, X_2 = 3, \text{ e } X_3 = 2.$$

A seguir determinaremos a densidade conjunta de $X_1, \ldots, X_r$. Para isso, sejam $x_1, x_2, \ldots, x_r$ r números inteiros não-negativos tais que $x_1 + \cdots + x_r = n$. Como as variáveis aleatórias $Y_1, Y_2, \ldots, Y_n$ são independentes e têm densidade comum, um pouco de raciocínio mostra que cada escolha específica x_1 delas tendo valor 1, x_2 delas tendo valor $2, \ldots, x_r$ delas tendo valor r tem a mesma probabilidade que é

$$p_1^{x_1} p_2^{x_2} \cdots p_r^{x_r}.$$

Assim, representando por $C(n; x_1, \ldots, x_r)$ o número de escolhas possíveis, vemos que

$$P(X_1 = x_1, \ldots, X_r = x_r) = C(n; x_1, \ldots, x_r) p_1^{x_1} \cdots p_r^{x_r}.$$

A determinação de $C(n; x_1, \ldots, x_r)$ é um problema de análise combinatória que se pode resolver facilmente pelos métodos do Capítulo 2. A maneira mais simples de fazê-lo é pensar nos r valores $1, 2, \ldots, r$ como r caixas e nas n provas como n bolas. Então $C(n; x_1, \ldots, x_r)$ é o número de maneiras em que podemos distribuir n bolas em r caixas de modo que tenhamos exatamente x_1, bolas na caixa $1, \ldots$, exatamente x_r bolas na caixa r. Se assim for feito, a caixa 1 terá x_1 bolas. Estas x_1 bolas poderão ser escolhidas dentre n bolas de

$\binom{n}{x_1}$ maneiras. As $n - x_1$ bolas restantes estarão distribuídas nas $r - 1$ caixas $2, 3, \ldots, r$ de modo que teremos x_2 bolas na caixa $2, \ldots, x_r$ bolas na caixa r. Assim

(16) $$C(n; x_1, \ldots, x_r) = \binom{n}{x_1} C(n - x_1; x_2, \ldots, x_r).$$

Segue-se por indução sobre n que

(17) $$C(n; x_1, \ldots, x_r) = \frac{n!}{(x_1!)(x_2!)\cdots(x_r!)}.$$

Realmente, para $r = 1$ não há nada a demonstrar. Suponha que (17) seja verdadeiro para $r - 1$ caixas. Então de (16) vemos que

$$\begin{aligned} C(n; x_1, \ldots, x_r) &= \frac{n!}{(x_1!)(n - x_1)!} \frac{(n - x_1)!}{(x_2!)\cdots(x_r!)} \\ &= \frac{n!}{(x_1!)\cdots(x_r!)} \end{aligned}$$

como desejado.

Portanto a densidade conjunta f de $X_1, \ldots, X_r$ é dada por

(18) $$f(x_1, \ldots, x_r) = \begin{cases} \dfrac{n!}{(x_1!)\cdots(x_r!)} p_1^{x_1} \cdots p_r^{x_r}, \\ x_i \text{ inteiros } \geqslant 0 \text{ tais que } x_1 + \cdots + x_r = n, \\ 0, \quad \text{outros valores de } x_i \end{cases}$$

Esta densidade chama-se *densidade multinomial* de parâmetros n e $p_1 \ldots, p_r$.

Observamos de imediato que as variáveis aleatórias $X_1, \ldots, X_r$ *não* são independentes. De fato, como $X_1 + X_2 + \cdots + X_r = n$, quaisquer $r - 1$ delas determinam a r-ésima. Usa-se às vezes esse fato, juntamente com a relação $p_1 + \cdots + p_r = 1$, para expressar a distribuição multinomial de uma forma diferente. Sejam $x_1, x_2, \ldots, x_{r-1}$; $r - 1$ números inteiros não-negativos tais que $x_1 + \cdots + x_{r-1} \leqslant n$. Então;

(19) $$\begin{aligned} &P(X_1 = x_1, \ldots, X_{r-1} = x_{r-1}) \\ &= \frac{n!}{(x_1!)\cdots(x_{r-1}!)(n - x_1 - \cdots - x_{r-1})!} \\ &\quad \times p_1^{x_1} \cdots p_{r-1}^{x_{r-1}} (1 - p_1 - \cdots - p_{r-1})^{n - x_1 - \cdots - x_{r-1}}. \end{aligned}$$

Esta forma é conveniente quando estamos interessados nos primeiros $r - 1$ resultados e pensamos no r-ésimo resultado como o resultado que é "nenhum dos

$r - 1$ resultados". Assim, no lançamento de um dado poderíamos estar interessados em saber se ocorrem 2, 4 ou 6. Então o experimento teria quatro resultados possíveis "2", "4", "6" e "não é (2, 4 ou 6)".

Seja k um número inteiro não negativo, $k \leq r$. Um simples argumento probabilístico mostra que para números inteiros não-negativos $x_1, x_2, \ldots, x_k$ tais que $x_1 + \cdots + x_k \leq n$,

$$(20) \quad P(X_1 = x_1, \ldots, X_k = x_k) = \frac{n!}{(x_1!) \cdots (x_k!)[n - (x_1 + \cdots + x_k)]!} \times p_1^{x_1} \cdots p_k^{x_k}(1 - (p_1 + \cdots + p_k))^{n-(x_1 + \cdots + x_k)}.$$

Para ver isto, observe que na realização das n provas estamos agora interessados somente nos $(k + 1)$ resultados "1", "2", ..., "k" e "não $(1, 2, \ldots, k)$". Assim, em essência temos n repetições de um experimento tendo $k + 1$ resultados, com X_i sendo o número de vezes que ocorre o i-ésimo resultado, $i = 1, 2, \ldots, k$. A equação (20) decorre então de (19) com $r - 1 = k$.

3.4.2 APROXIMAÇÃO DE POISSON PARA DISTRIBUIÇÃO BINOMIAL. Existe uma importante ligação entre a distribuição binomial e a de Poisson. Suponha por exemplo que realizamos n provas de Bernoulli com probabilidade de sucesso $p_n = \lambda/n$ em cada prova. Então a probabilidade de obter $S_n = k$ sucessos nas n provas é dada por

$$P(S_n = k) = \binom{n}{k} (p_n)^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} \frac{(n)_k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

a medida que $n \to \infty$, $(n)_k/n^k \to 1$, $(1 - \lambda/n)^n \to e^{-\lambda}$, e $(1 - \lambda/n)^{-k} \to 1$.

Conseqüentemente,

$$(21) \quad \lim_{n \to \infty} \binom{n}{k} (p_n)^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Na derivação de (21) tínhamos $np_n = \lambda$. Na realidade (21) é válido sempre que $np_n \to \lambda$ quando $n \to \infty$.

Usa-se a Equação (21) nas aplicações para aproximar a distribuição binomial pela distribuição de Poisson quando a probabilidade de sucesso p é pequena e n é grande. Faz-se isso aproximando a probabilidade binomial $P(S_n = x)$ por meio de $f(x)$, onde f é a densidade de Poisson de parâmetro $\lambda = np$. A aproximação é bastante quando np^2 é pequeno. O exemplo seguinte ilustra essa técnica.

Exemplo 15. Uma máquina produz parafusos dos quais 19 são defeituosos. Determine a probabilidade de que não haja parafusos defeituosos numa caixa de 200 parafusos.

Temos aqui $n = 200$ provas de Bernoulli com probabilidade de sucesso $p = 0{,}01$. A probabilidade de que não haja parafusos defeituosos é

$$(1 - 0{,}01)^{200} = (0{,}99)^{200} = 0{,}1340.$$

A aproximação de Poisson para esta probabilidade é dada por

$$e^{-200(0{,}01)} = e^{-2} = 0{,}1353.$$

O fato de que a distribuição de Poisson pode ocorrer como o limite de distribuições binomiais tem importantes conseqüências teóricas. É uma das justificativas para desenvolver modelos baseados em processo de Poisson, que serão discutidos no Capítulo 9. O uso da aproximação de Poisson como artifício para poupar trabalho na determinação de probabilidade binomial tem importância secundária, uma vez que se pode calcular facilmente as próprias probabilidades binomiais.

3.5. SEQÜÊNCIAS INFINITAS DE PROVAS DE BERNOULLI

Considere a realização repetida de um experimento do tipo sucesso-fracasso com probabilidade de sucesso p até que ocorra o primeiro sucesso. Para qualquer número predeterminado n de provas, existe a probabilidade não-nula $(1 - p)^n$ de que não ocorra nenhum sucesso. Assim, considerando o número de provas até o primeiro sucesso, não podemos nos limitar a qualquer número predeterminado de provas mas, considerar uma seqüência interminável de provas.

Um dado número finito n de provas constituem n provas de Bernoulli, representadas por n variáveis aleatórias independentes de Bernoulli $X_1, \ldots, X_n$. Para representar uma seqüência infinita de provas de Bernoulli consideramos uma seqüência infinita $\{X_n\}$, $n \geqslant 1$, de variáveis aleatórias independentes de Bernoulli tendo o mesmo parâmetro p.

Em geral diz-se que as variáveis aleatórias $X_1, X_2, \ldots$ são independentes se as variáveis aleatórias $X_1, \ldots, X_n$ forem mutuamente independentes para qualquer número inteiro positivo n. Pode-se mostrar que dada uma densidade discreta qualquer f, existe um espaço de probabilidade $(\Omega, \mathscr{A}, P)$ no qual estão definidas as variáveis aleatórias mutuamente independentes $X_1, X_2, \ldots$ com densidade comum f.

Como modelo para a realização de uma seqüência ilimitada de Bernoulli tomamos, portanto, uma seqüência infinita $\{X_n\}$, $n \geqslant 1$, de variáveis aleatórias de Bernoulli mutuamente independentes tais que $P(X_n = 1) = p$, $n \geqslant 1$. Interpretamos $X_n = 1$ como significando que a n-ésima prova resulta em sucesso e $X_n = 0$ significando que ela resulta em fracasso.

Considere o número de provas W_1 até o primeiro sucesso. A variável aleatória W_1 pode assumir somente os valores inteiros $1, 2, \ldots$ O evento $\{W_1 = n\}$ ocorre se, e somente se, as primeiras $n - 1$ provas são fracassos e a n-ésima prova é um sucesso. Portanto

$$\{W_1 = n\} = \{X_1 = 0, \ldots, X_{n-1} = 0, X_n = 1\}.$$

Segue-se que

$$\begin{aligned} P(W_1 = n) &= P(X_1 = 0, \ldots, X_{n-1} = 0, X_n = 1) \\ &= P(X_1 = 0) \cdots P(X_{n-1} = 0) P(X_n = 1) \\ &= (1-p)^{n-1} p. \end{aligned}$$

Conseqüentemente

(22) $$P(W_1 - 1 = n) = p(1-p)^n.$$

Assim, $W_1 - 1$ distribui-se geometricamente com parâmetro p.

Seja $r \geqslant 1$ um número inteiro e T_r o número de provas até o r-ésimo sucesso (de modo que r-ésimo sucesso ocorre na prova T_r). Então T_r é uma variável aleatória que pode assumir somente os valores $r, r+1, \ldots$ O evento $\{T_r = n\}$ ocorre se, e somente se, ocorrer um sucesso na n-ésima prova e ocorrem exatamente $r - 1$ sucessos nas primeiras $n - 1$ provas. Assim

$$\{T_r = n\} = \{X_1 + \cdots + X_{n-1} = r - 1\} \cap \{X_n = 1\}.$$

Como os dois eventos do segundo membro são independentes e $X_1 + \cdots + X_{n-1}$ tem distribuição binomial de parâmetros $n - 1$ e p, vemos que para $n = r, r+1, \ldots$

$$\begin{aligned} P(T_r = n) &= P(X_1 + \cdots + X_{n-1} = r - 1) P(X_n = 1) \\ &= \binom{n-1}{r-1} p^{r-1}(1-p)^{n-r} p \\ &= \binom{n-1}{r-1} p^r (1-p)^{n-r}. \end{aligned}$$

Conseqüentemente

(23) $$P(T_r - r = n) = \binom{r+n-1}{r-1} p^r (1-p)^n.$$

Vemos das equações (4) e (23) que $T_r - r$ tem distribuição binomial negativa de parâmetros $\alpha = r$ e p.

Seja $T_0 = 0$ e T_r a variável aleatória definida acima para qualquer número inteiro $r \geqslant 1$. Defina $W_i = T_i - T_{i-1}, i = 1, 2, \ldots$ Então W_i é o número de provas

até o i-ésimo sucesso, após o $(i-1)$ i-ésimo sucesso. Mostraremos a seguir que, para qualquer número inteiro $r \geqslant 1$, as variáveis aleatórias $W_1 - 1, W_2 - 1, \ldots, W_r - 1$ são mutuamente independentes e têm a mesma densidade geométrica de parâmetro p.

Para tanto, sejam $n_1, n_2, \ldots, n_r$ r números inteiros positivos quaisquer. Então o evento $\{W_1 = n_1, \ldots, W_r = n_r\}$ ocorre somente se todas as primeiras $n_1 + \cdots + n_r$ provas são fracassos, exceto as provas

$$n_1, n_1 + n_2, \ldots, n_1 + \cdots + n_r,$$

que são sucessos. Como as provas são mutuamente independentes com probabilidade comum de sucesso p, vemos que

$$P(W_1 = n_1, \ldots, W_r = n_r) = (1-p)^{n_1 - 1} p (1-p)^{n_2 - 1} p \cdots (1-p)^{n_r - 1} p$$

$$= \prod_{i=1}^{r} [p(1-p)^{n_i - 1}].$$

Assim as variáveis aleatórias $W_1 - 1, \ldots, W_r - 1$ são independentes e distribuem-se geometricamente com parâmetro p.

Vemos claramente que $T_r - r = (W_1 - 1) + \cdots + (W_r - 1)$, de modo que $T_r - r$ é a soma de r variáveis aleatórias independentes geometricamente distribuídas. Mostramos anteriormente que $T_r - r$ tem distribuição binomial negativa de parâmetros r e p. Fica assim estabelecido o fato interessante e importante de que *a distribuição da soma de* r *variáveis aleatórias geométricas independentes identicamente distribuídas com parâmetro* p *tem distribuição binomial negativa de parâmetros* r *e* p.

Propriedades adicionais de seqüências infinitas de provas de Bernoulli serão tratadas nos exercícios.

3.6. SOMA DE VARIÁVEIS ALEATÓRIAS INDEPENDENTES

Nesta seção discutiremos métodos para determinar a distribuição da soma de um número finito de variáveis aleatórias discretas independentes. Começamos considerando duas de tais variáveis aleatórias, digamos X e Y.

Suponhamos então que X e Y são variáveis aleatórias discretas independentes. Sejam $x_1, x_2, \ldots$ os distintos valores possíveis de X. Para qualquer z, o evento $\{X + Y = z\}$ é o mesmo que o evento

$$\bigcup_i \{X = x_i, Y = z - x_i\}.$$

Como os eventos $\{X = x_i, Y = z - x_i\}$ são disjuntos para valores distintos de i, segue-se que

$$P(X + Y = z) = \sum_i P(X = x_i, Y = z - x_i)$$

$$= \sum_i P(X = x_i)P(Y = z - x_i) = \sum_i f_X(x_i) f_Y(z - x_i).$$

Em outras palavras,

(24) $$f_{X+Y}(z) = \sum_x f_X(x) f_Y(z - x).$$

Se X e Y são variáveis aleatórias inteiras, então $X + Y$ é também inteira. Neste caso podemos interpretar (24) como sendo válida, quando z é inteiro e a variável x do segundo membro se estende sobre os (24) inteiros. Uma particularização adicional é útil. Suponha que X e Y assumam somente valores inteiros não-negativos. Então $X + Y$ também assume somente valores inteiros não-negativos. Se z é um número inteiro não-negativo, então $f_X(x)\, f_Y(z - x) = 0$ a menos que x seja um dos valores $0, 1, \ldots, z$. Então sob essas condições podemos reescrever (24) como

(25) $$f_{X+Y}(z) = \sum_{x=0}^{z} f_X(x) f_Y(z - x).$$

Embora a equação (25) seja útil na determinação da densidade de $X + Y$, geralmente é mais simples usar funções geratrizes de probabilidade. Descreveremos a seguir tais funções e apresentaremos diversas aplicações importantes de seu uso na determinação da densidade de soma de variáveis aleatórias independentes.

Definição 6. Seja X uma variável aleatória inteira não-negativa. Define-se função geratriz de probabilidade Φ_X de X como

$$\Phi_X(t) = \sum_{x=0}^{n} P(X = x)t^x = \sum_{x=0}^{n} \binom{n}{x} (pt)^x (1 - p)^{n-x}.$$

A seguir determinaremos $\Phi_X(t)$ para três casos específicos.

Exemplo 16. Distribuição binomial. Seja X uma variável aleatória tendo distribuição binomial de parâmetros n e p. Então

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

e portanto

$$\Phi_X(t) = \sum_{x=0}^{\infty} P(X = x)t^x = \sum_{x=0}^{\infty} f_X(x)t^x, \qquad -1 \le t \le 1.$$

Pela fórmula de expansão binomial

$$(a + b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x},$$

concluímos que

(26) $$\Phi_X(t) = (pt + 1 - p)^n.$$

Exemplo 17. Distribuição binomial negativa. Seja X uma variável aleatória com distribuição binomial negativa de parâmetros α e p. Então

$$P(X = x) = p^\alpha \binom{-\alpha}{x} (-1)^x (1 - p)^x$$

e portanto

$$\Phi_X(t) = \sum_{x=0}^{\infty} p^{\alpha} \binom{-\alpha}{x} (-1)^x (1-p)^x t^x$$

$$= p^{\alpha} \sum_{x=0}^{\infty} \binom{-\alpha}{x} (-t(1-p))^x.$$

Da expansão em série de Taylor

$$(1+s)^{-\alpha} = \sum_{x=0}^{\infty} \binom{-\alpha}{x} s^x,$$

com $s = -t(1-p)$, segue-se que

$$\Phi_X(t) = \left(\frac{p}{1-t(1-p)}\right)^{\alpha}. \tag{27}$$

Exemplo 18. Distribuição de Poisson. Suponha que X tem uma distribuição de Poisson de parâmetro λ. Então

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

e portanto

$$\Phi_X(t) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda t)^x}{x!}.$$

Fazendo $s = \lambda t$ na expansão em série de Taylor

$$e^s = \sum_{x=0}^{\infty} \frac{s^x}{x!},$$

vemos que

$$\Phi_X(t) = e^{\lambda t} e^{-\lambda} = e^{\lambda(t-1)}. \tag{28}$$

Sejam X e Y duas variáveis aleatórias independentes, inteiras e não-negativas. Então

$$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t). \tag{29}$$

Para verificar este resultado, observe que em virtude de (25)

$$\Phi_{X+Y}(t) = \sum_{z=0}^{\infty} f_Z(z) t^z$$

$$= \sum_{z=0}^{\infty} t^z \sum_{x=0}^{z} f_X(x) f_Y(z-x)$$

$$= \sum_{x=0}^{\infty} f_X(x) t^x \sum_{z=x}^{\infty} f_Y(z-x) t^{z-x}$$

$$= \sum_{x=0}^{\infty} f_X(x) t^x \sum_{y=0}^{\infty} f_Y(y) t^y$$

$$= \Phi_X(t)\Phi_Y(t),$$

que é o resultado desejado.

De (29) segue-se facilmente por indução que se $X_1, \ldots, X_r$ são variáveis aleatórias independentes, inteiras e não-negativas, então

(30) $$\Phi_{X_1+\cdots+X_r}(t) = \Phi_{X_1}(t) \cdots \Phi_{X_r}(t).$$

Pode-se demonstrar mais facilmente as conclusões do teorema seguinte usando a "técnica da função geratriz", que se baseia no fato de que se

$$\sum_{x=0}^{\infty} a_x t^x = \sum_{x=0}^{\infty} b_x t^x, \qquad -1 < t < 1,$$

então podemos igualar os coeficientes de t^x nas duas séries de potência e concluir que $a_x = b_x$, $x = 0, 1, 2, \ldots$ Isto mostra que se duas variáveis aleatórias inteiras e não-negativas têm a mesma função geratriz de probabilidade, elas devem ter a mesma distribuição. Em outras palavras, a função geratriz de probabilidade de uma variável aleatória inteira não-negativa determina unicamente a distribuição desta variável aleatória.

Teorema 1. Sejam $X_1, \ldots, X_r$ variáveis aleatórias independentes.

(i) Se X_i tem distribuição binomial de parâmetros n_i e p, então $X_1 + \cdots + X_r$ tem distribuição binomial de parâmetros $n_1 + \cdots + n_r$ e p.

(ii) Se X_i tem distribuição binomial negativa de parâmetros α_i e p, então $X_1 + \cdots + X_r$ tem distribuição binomial negativa de parâmetros $\alpha_1 + \cdots + \alpha_r$ e p.

(iii) Se X_i tem distribuição de Poisson de parâmetro λ_i, então $X_1 + \cdots + X_r$ tem distribuição de Poisson de parâmetro $\lambda_1 + \cdots + \lambda_r$.

Demonstração de (i). Se os X_i são como especificados em (i), então pelo Exemplo (16)

$$\begin{aligned}\Phi_{X_1+\cdots+X_r}(t) &= \Phi_{X_1}(t) \cdots \Phi_{X_r}(t) \\ &= (pt + 1 - p)^{n_1} \cdots (pt + 1 - p)^{n_r} \\ &= (pt + 1 - p)^{n_1+\cdots+n_r}.\end{aligned}$$

Assim a função geratriz de probabilidade de $X_1 + \cdots + X_r$ é igual à de uma variável aleatória com distribuição binomial de parâmetros $n_1 + \cdots + n_r$ e p. Isto implica que $X_1 + \cdots + X_r$ deve ter esta distribuição binomial. Pois suponha que

$$a_x = \binom{n_1 + \cdots + n_r}{x} p^x (1 - p)^{n_1+\cdots+n_r-x}$$

sejam as probabilidades binomiais correspondentes. Então

$$\begin{aligned}\sum_{x=0}^{\infty} P(X_1 + \cdots + X_r = x)t^x &= \Phi_{X_1+\cdots+X_r}(t) \\ &= (pt + 1 - p)^{n_1+\cdots+n_r} \\ &= \sum_{x=0}^{\infty} a_x t^x.\end{aligned}$$

Assim, igualando os coeficientes, vemos que

$$P(X_1 + \cdots + X_r = x) = a_x$$

e portanto que $X_1 + \cdots + X_r$ distribui-se binomialmente conforme estabelece (i).

Demonstração de (ii). Se os X_i são como especificados em (ii), então pelo Exemplo 17.

$$\begin{aligned}\Phi_{X_1+\cdots+X_r}(t) &= \Phi_{X_1}(t)\cdots\Phi_{X_r}(t)\\ &= \left(\frac{p}{1-t(1-p)}\right)^{\alpha_1}\cdots\left(\frac{p}{1-t(1-p)}\right)^{\alpha_r}\\ &= \left(\frac{p}{1-t(1-p)}\right)^{\alpha_1+\cdots+\alpha_r}\end{aligned}$$

Assim a função geratriz de probabilidade de $X_i + \cdots + X_r$ é igual à de uma variável aleatória com distribuição binomial negativa de parâmetros $\alpha_1 + \cdots + \alpha_r$ e p. Segue-se então pelo mesmo argumento usado na demonstração de (i) que $X_1 + \cdots + X_r$ tem a distribuição binomial negativa especificada.

A demonstração de (iii) é análoga as de (i) e (ii), e deixamo-la para o leitor a título de exercício.

Suponha que $\alpha_1 = \cdots = \alpha_r = 1$ no item (ii) do Teorema 1. Então cada uma das variáveis aleatórias $X_1, \ldots, X_r$ distribui-se geometricamente com parâmetro p, e (ii) estabelece que $X_1 + \cdots + X_r$ tem distribuição binomial negativa de parâmetros r e p. Isto constitui uma demonstração alternativa do resultado obtido na Seção 3.5.

O exemplo seguinte ilustra o uso de probabilidades condicionais.

Exemplo 19. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, inteiras e não-negativas, tendo uma densidade comum. Suponha que $S_0 = 0$ e $S_n = X_1 + \cdots + X_n$, $n \geqslant 1$. Seja N uma variável aleatória inteira não-negativa e suponha que $N, X_1, X_2, \ldots$ são independentes. Então $S_N = X_1 + \cdots + X_N$ é a soma de um número aleatório de variáveis aleatórias. Para uma interpretação de S_N, suponha que no tempo 0 um número aleatório N de bactérias são introduzidos num sistema, e que no tempo 1 a colônia iniciada pela i-ésima bactéria contém X_i membros. Então S_N é o número total de bactérias presentes no tempo 1. Mostre que a função geratriz de probabilidade de S_N é dada por

$$\Phi_{S_N}(t) = \Phi_N(\Phi_{X_1}(t)), \quad -1 \leqslant t \leqslant 1. \tag{31}$$

Para verificar (31) observamos inicialmente que

$$P(S_N = x) = \sum_{n=0}^{\infty} P(S_N = x, N = n)$$

$$= \sum_{n=0}^{\infty} P(S_n = x, N = n)$$

$$= \sum_{n=0}^{\infty} P(N = n)P(S_n = x \mid N = n).$$

Como N é independente de $X_1, X_2, \ldots, X_n$, é independente de S_n, e portanto $P(S_n = x/N = n) = P(S_n = x)$. Assim

$$P(S_N = x) = \sum_{n=0}^{\infty} P(N = n)P(S_n = x). \tag{32}$$

Conseqüentemente para $-1 \leqslant t \leqslant 1$

$$\Phi_{S_N}(t) = \sum_{x=0}^{\infty} t^x P(S_N = x)$$

$$= \sum_{x=0}^{\infty} t^x \sum_{n=0}^{\infty} P(N = n)P(S_n = x)$$

$$= \sum_{n=0}^{\infty} P(N = n) \sum_{x=0}^{\infty} t^x P(S_n = x)$$

$$= \sum_{n=0}^{\infty} P(N = n)\, \Phi_{S_n}(t)$$

$$= \sum_{n=0}^{\infty} P(N = n)(\Phi_{X_1}(t))^n = \Phi_N(\Phi_{X_1}(t)).$$

Exercícios

1. Qualquer ponto no intervalo $[0, 1)$ pode ser representado através de sua expansão decimal $0, x_1, x_2 \ldots$ Suponha que se escolhe, aleatoriamente, um ponto do intervalo $[0, 1)$. Seja X o primeiro dígito da expansão decimal que representa o ponto. Determine a densidade de X.
2. Suponha que X tenha uma densidade binomial negativa de parâmetros $\alpha = r$ e p, (sendo r inteiro). Determine a densidade de $X + r$.
3. Suponha que uma caixa contenha 6 bolas vermelhas e 4 pretas. Seleciona-se uma amostra aleatória de tamanho n. Seja X o número de bolas vermelhas na amostra. Determine a densidade de X para amostragem (a) sem reposição e (b) com reposição.
4. Seja N um número inteiro positivo e seja

$$f(x) = \begin{cases} c2^x, & x = 1, 2, \ldots, N, \\ 0, & \text{para outros valores de } x. \end{cases}$$

Determine o valor de c para o qual f é uma densidade de probabilidade.

5. Suponha que X é uma variável aleatória tendo a densidade f dada por

x	−3	−1	0	1	2	3	5	8
$f(x)$	0,1	0,2	0,15	0,2	0,1	0,15	0,05	0,05

Determine as seguintes probabilidades:

(a) X é negativo;
(b) X é par;
(c) X assume um valor entre 1 e 8 (inclusivos);
(d) $P(X = -3/X \leqslant 0)$
(e) $P(X \geqslant 3/X > 0)$.

6. Suponha que X tem uma distribuição geométrica com $p = 0{,}8$.
Determine as probabilidades dos seguintes eventos:

(a) $X > 3$;
(b) $4 \leqslant X \leqslant 7$ ou $X > 9$;
(c) $3 \leqslant X \leqslant 5$ ou $7 \leqslant X \leqslant 10$.

7. Suponha que X se distribui uniformemente sobre $0, 1, \ldots, 99$. Determine:

(a) $P(X \geqslant 25)$;
(b) $P(2{,}6 < X < 12{,}2)$;
(c) $P(8 < X \leqslant 10$ ou $30 < X \leqslant 32)$;
(d) $P(25 \leqslant X \leqslant 30)$.

8. Suponha que uma caixa contém 12 bolas numeradas de 1 a 12. Faz-se duas repetições independentes do experimento de selecionar aleatoriamente uma bola da caixa. Seja X o maior entre os dois números observados. Determine a densidade de X.

9. Considere a situação do Exercício 8 em que a seleção é feita sem reposição. Determine a densidade de X.

10. Seja X uma variável aleatória geometricamente distribuída com parâmetro p. Seja $Y = X$ se $X < M$ e $Y = M$ se $X \geqslant M$; isto é, $Y = \text{Min}(X, M)$. Determine a densidade de Y.

11. Suponha que X se distribui geometricamente com parâmetro p. Determine a densidade de

(a) X^2;
(b) $X + 3$.

12. Suponha que uma caixa contém r bolas numeradas de 1 a r. Seleciona-se sem reposição uma amostra aleatória de tamanho n. Seja Y o maior número observado na amostra e Z o menor.

(a) Determine a probabilidade $P(Y \leqslant y)$.
(b) Determine a probabilidade $P(Z \geqslant z)$.

13. Sejam X e Y duas variáveis aleatórias tendo a densidade conjunta dada na tabela seguinte.

X \ Y	-1	0	2	6
-2	1/9	1/27	1/27	1/9
1	2/9	0	1/9	1/9
3	0	0	1/9	4/27

Determine as probabilidades dos seguintes eventos:

(a) Y é par;
(b) XY é ímpar;
(c) $X > 0$ e $Y \geqslant 0$.

14. Seja X e Y duas variáveis aleatórias independentes com densidade uniforme em $\{0, 1, \ldots, N\}$. Determine

(a) $P(X \geqslant Y)$;
(b) $P(X = Y)$.

15. Sejam X e Y como no Exercício 14. Obtenha as densidades de:

(a) $\min(X, Y)$;
(b) $\max(X, Y)$;
(c) $|Y - X|$.

16. Sejam X e Y duas variáveis aleatórias independentes tendo densidades geométricas de parâmetros p_1 e p_2 respectivamente.
Obtenha:

(a) $P(X \geqslant Y)$;
(b) $P(X = Y)$.

17. Sejam X e Y como no Exercício 16. Obtenha a densidade de

(a) $\min(X, Y)$;
(b) $X + Y$.

18. Sejam X e Y variáveis aleatórias discretas e sejam g e h funções tais que satisfaçam a identidade.

$$P(X = x, Y = y) = g(x)h(y).$$

(a) Expresse $P(X = x)$ em termos de g e h.
(b) Expresse $P(Y = y)$ em termos de g e h.
(c) Mostre que $(\Sigma_x\, g(x))\, (\Sigma_y\, h(y)) = 1$.
(d) Mostre que X e Y são independentes.

19. Sejam X e Y variáveis aleatórias independentes com a mesma densidade geométrica de parâmetro p. Sejam $Z = Y - X$ e $M = \min(X, Y)$.

(a) Mostre que para z e $m \geqslant 0$ inteiros

$$P(M = m, Z = z) = \begin{cases} P(X = m - z)P(Y = m), & z < 0, \\ P(X = m)P(Y = m + z), & z \geqslant 0. \end{cases}$$

(b) Conclua de (a) que para z e $m \geqslant 0$ inteiros

$$P(M = m, Z = z) = p^2(1 - p)^{2m}(1 - p)^{|z|}.$$

(c) Use (b) e o Exercício 18 para mostrar que M e Z são independentes.

20. Suponha que um alvo circular está dividido em três zonas limitadas pelos círculos concêntricos de raios 1/3, 1/2 e 1, conforme ilustra o diagrama seguinte.

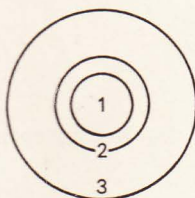



Figura 4

Disparando-se aleatoriamente três tiros ao alvo, qual a probabilidade de que exatamente um tiro atinja cada zona?

21. Suponha que se distribui aleatoriamente $2r$ bolas em r caixas. Seja X_i o número de bolas na caixa i.

(a) Obtenha a densidade conjunta de $X_1, \ldots, X_r$.

(b) Obtenha a probabilidade de que cada caixa contenha exatamente 2 bolas.

22. Considere um experimento com três resultados possíveis que ocorrem com probabilidades p_1, p_2 e p_3, respectivamente. Suponha que se realiza n repetições independentes do experimento e seja X_i o número de vezes que ocorre o resultado i.

(a) Qual é a densidade de $X_1 + X_2$?

(b) Determine $P(X_2 = y \mid X_1 + X_2 = z)$, $y = 0, 1, 2, \ldots, z$.

23. Use a aproximação de Poisson para calcular a probabilidade de que no máximo 2 dentre 50 motoristas tenham carteiras de habilitação inválida se normalmente 5% dos motoristas o tem.

24. Use a aproximação de Poisson para calcular a probabilidade de que uma caixa com 100 fusíveis contenha no máximo 2 fusíveis defeituosos se 3% dos fusíveis fabricados são defeituosos.

25. Lança-se um dado até observar o número 6.

(a) Qual é a probabilidade de que sejam necessários seis lançamentos no máximo?

(b) Quantos lançamentos são necessários para que a probabilidade de obter 6 seja no mínimo 1/2?

Os Exercícios 26-30 são problemas inter-relacionados e baseiam-se em uma seqüência infinita de provas de Bernoulli conforme discussão da Seção 3.5.

26. Seja T_i o número de provas até (e inclusive) o i-ésimo sucesso. Sejam $0 \leqslant x_1 < \cdots < x_r$ números inteiros. Determine a probabilidade

$$P(T_1 = x_1, T_2 = x_2, \ldots, T_r = x_r).$$

Sugestão: Faça $W_r = T_r - T_{r-1}, r \geqslant 2$ e $W_1 = T_1$; então

$$P(T_1 = x_1, \ldots, T_r = x_r)$$
$$= P(W_1 = x_1, W_2 = x_2 - x_1, \ldots, W_r = x_r - x_{r-1}).$$

Use a seguir o fato de que as variáveis aleatórias $W_1 - 1, \ldots, W_r - 1$ são mutuamente independentes e têm a mesma distribuição geométrica de parâmetro p.

27. Seja N_n o número de sucessos nas primeiras n provas. Mostre que

$$P(T_1 = x \mid N_n = 1) \frac{1}{n}, \qquad x = 1, 2, \ldots, n.$$

28. De uma forma mais geral, mostre que

$$P(T_1 = x_1, T_2 = x_2, \ldots, T_r = x_r \mid N_n = r) = \binom{n}{r}^{-1},$$
$$0 < x_1 < x_2 < \cdots < x_r \leqslant n.$$

Isto mostra que, dado que ocorrem r sucessos nas primeiras n provas, as provas nas quais tais sucessos ocorrem constituem uma amostra aleatória de tamanho r (sem reposição) da "população" das localizações possíveis.

29. Seja k um número inteiro positivo, $k \leqslant r$. Do Exercício 28 podemos determinar de imediato que

$$P(T_k = x \mid N_n = r) = \frac{\binom{x-1}{k-1}\binom{n-x}{r-k}}{\binom{n}{r}}.$$

Realmente, se $T_k = x$, o k-ésimo sucesso ocorre na posição x. Nas primeiras $x - 1$ posições deve ocorrer exatamente $k - 1$ sucessos, e nas últimas $n - x$ posições deve ocorrer exatamente $r - k$ sucessos. Como, dado $N_n = r$, as posições dos r sucessos constituem uma amostra aleatória de tamanho r da "população" de n posições, segue-se o resultado acima. Verifique que isto realmente acontece determinando $P(T_k = x/N_n = r)$ diretamente.

30. Sejam $1 \leqslant i < j \leqslant r$ números inteiros não-negativos. Obtenha

$$P(T_i = x, T_j = y \mid N_n = r)$$

para $0 < x < y \leqslant n$.

31. Suponha que X e Y sejam variáveis aleatórias independentes que tenham densidade uniforme sobre $1, 2, \ldots, N$. Determine a densidade de $X + Y$.

32. Seja X uma variável aleatória uniformemente distribuída em $\{0, 1, 2, \ldots N\}$. Obtenha $\Phi_X(t)$.

33. Seja X uma variável aleatória inteira não-negativa cuja função geratriz de probabilidade é dada por $\Phi_X(t) = e^{\lambda(t^2 - 1)}$, onde $\lambda > 0$. Obtenha f_X.

34. Demonstre o item (iii) do Teorema 1.

35. Sejam X e Y variáveis aleatórias independentes com densidades de Poisson de parâmetros λ_1 e λ_2, respectivamente. Obtenha $P(Y = y/X + Y = z)$ para $y = 0, \ldots, z$. *Sugestão:* use (iii) do Teorema 1.

36. Sejam X, Y, e Z variáveis aleatórias independentes com densidades de Poisson de parâmetros λ_1, λ_2 e λ_3, respectivamente. Obtenha

$$P(X = x, Y = y, Z = z \mid X + Y + Z = x + y + z)$$

para números inteiros não-negativos x, y e z. *Sugestão:* use (iii) do Teorema 1.

37. No Exemplo 19 suponha que X_1 assuma os valores 1 e 0 com probabilidades p e $1-p$, respectivamente, onde $0 < p < 1$. Suponha também que N tenha densidade de Poisson de parâmetro λ.

    (a) Use a Equação (31) para obter a função geratriz de probabilidade de S_N.
    (b) Use (a) para obter a densidade de S_N.

    Para uma interpretação de S_N, suponha que se introduza um número aleatório N de células cancerosas no tempo 0 e que cada célula, independentemente das outras células e de N, tem probabilidade p de sobreviver a um tratamento radioativo. Seja $X_i = 1$ se a i-ésima célula sobrevive e $X_i = 0$ caso contrário. Então S_N é o número total de células que sobrevivem ao tratamento.

38. Resolva (b) do Exercício 37 sem usar a função geratriz de probabilidade, mas usando em seu lugar a Equação (32) e o fato de que $X_1 + \cdots + X_n$ tem uma densidade binomial.

4

EXPECTÂNCIA DE VARIÁVEIS ALEATÓRIAS INDEPENDENTES

Consideremos um certo jogo de azar. Para apostar no jogo devemos pagar uma importância de a dólares. Como resultado da aposta recebemos X dólares, onde X é uma variável aleatória cujos valores possíveis são $x_1, x_2, \ldots, x_r$. O problema é saber se devemos apostar no jogo. Se vamos apostar uma única vez, o problema é bastante difícil. Entretanto, suponha que apostamos um grande número de vezes. Após n apostas, pagaríamos na dólares e receberíamos $X_1 + \cdots + X_n$ dólares. Se supomos que as apostas sucessivas do jogo constituem repetições independentes do mesmo experimento (observação de um valor de X), podemos tomar as variáveis aleatórias $X_1, X_2, \ldots, X_n$ como sendo mutuamente independentes e tendo a densidade comum f de X. Seja $N_n(x_i)$ o número de apostas que resultam no valor x_i, isto é, o número de X_i que assumem o valor x_i. Então podemos escrever

$$X_1 + \cdots + X_n = \sum_{i=1}^{r} x_i N_n(x_i).$$

A quantia média recebida é então

$$\frac{X_1 + \cdots + X_n}{n} = \sum_{i=1}^{r} x_i \left[\frac{N_n(x_i)}{n}\right].$$

De acordo com a interpretação da probabilidade como freqüência relativa se n for grande, os números $N_n(x_i)/n$ devem ser aproximadamente iguais a $f(x_i)$, e assim a soma do segundo membro deve ser aproximadamente igual a $\mu = \Sigma_{i=1}^{r} x_i f(x_i)$. Assim parece razoável antecipar um ganho líquido no jogo se $\mu > a$ e esperar uma perda líquida se $\mu < a$. Se $\mu = a$, anteciparíamos sair do jogo mais ou menos quites.

A quantidade $\Sigma_{i=1}^{r} x_i f(x_i)$ chama-se expectância da variável aleatória X. De uma forma mais geral, seja X uma variável aleatória discreta qualquer, que assume um número finito de valores $x_1, \ldots, x_r$. Então o valor esperado de X, representado por EX ou μ, é o número

$$EX = \sum_{i=1}^{r} x_i f(x_i), \tag{1}$$

onde f é a densidade de X.

Suponha que X tem distribuição uniforme em $x_1, \ldots, x_r$. Então $f(x_i) = P(X = x_i) = r^{-1}$ e de (1) vemos que $EX = (x_1 + \cdots + x_r)r^{-1}$, de modo que neste caso, EX é apenas a média aritmética dos valores possíveis de X. Em geral (1) mostra que EX é a *média aritmética* dos valores possíveis de X; o peso associado ao i-ésimo valor x_i é sua probabilidade $f(x_i)$.

O valor esperado EX chama-se, também, média de X (ou da densidade f de X) e representa-se habitualmente por μ. A média é uma forma de tentar reduzir a distribuição de probabilidade a um único número que se supõe representar o "valor típico" de X. O grau de adequação disto depende do grau de concentração dos valores de X em torno de μ. Examinaremos este problema em maior detalhe quando discutirmos a variância de X na Seção 4.3.

Exemplo 1. Distribuição binomial. Suponha que X tem a distribuição binomial de parâmetros n e p. Obtenha EX.

Para $n = 1$, X assume os valores 0 e 1 com probabilidades $(1 - p)$ e p, respectivamente. Portanto:

$$EX = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = p.$$

Como uma variável aleatória, que tem densidade binomial de parâmetros 1 e p, é simplesmente uma variável aleatória indicadora, vemos que se pode obter a probabilidade de um evento A, de que $X = 1$, determinando a expectância do seu indicador.

Determinemos agora EX para $n \geqslant 1$ qualquer. Neste caso X assume os valores $0, 1, 2, \ldots, n$, e

$$EX = \sum_{j=0}^{n} j \binom{n}{j} p^j (1 - p)^{n-j}.$$

Para calcular esta quantidade observamos que

$$\begin{aligned} j\binom{n}{j} &= \frac{jn!}{j!(n - j)!} \\ &= \frac{n(n - 1)!}{(j - 1)![(n - 1) - (j - 1)]!} \\ &= n\binom{n - 1}{j - 1}. \end{aligned}$$

Assim

$$EX = n \sum_{j=1}^{n} \binom{n - 1}{j - 1} p^j (1 - p)^{n-j}.$$

Fazendo a mudança de variável $i = j - 1$ vemos que

$$EX = np \sum_{i=0}^{n-1} \binom{n - 1}{i} p^i (1 - p)^{n-i-1}.$$

Pelo teorema binomial

$$\sum_{i=0}^{n-1} \binom{n-1}{i} p^i(1-p)^{n-i-1} = [p + (1-p)]^{n-1} = 1,$$

de modo que

$$EX = np.$$

4.1. DEFINIÇÃO DE EXPECTÂNCIA

Suponha agora que X é uma variável aleatória discreta qualquer, cujos valores possíveis são $x_1, x_2, \ldots$. Gostaríamos de definir a expectância de X como

$$EX = \sum_{j=1}^{\infty} x_j f(x_j). \tag{2}$$

Se X tem apenas um número finito de valores possíveis $x_1, \ldots, x_r$, então (2) é simplesmente a nossa definição anterior. No caso discreto geral, esta definição é válida desde que a soma $\Sigma_j\, x_j\, f(x_j)$ seja bem definida. Para que este seja o caso, exigimos que $\Sigma_j\, |x_j|\, f(x_j) < \infty$. Isto nos leva à seguinte definição.

Definição 1. Seja X uma variável aleatória discreta com densidade f. Se $\Sigma_j\, |x_j|\, f(x_j) < \infty$, dizemos que X tem expectância finita e definimos sua expectância através de (2). Por outro lado, se $\Sigma_{j=1}^{\infty}\, |x_j|\, f(x_j) = \infty$, dizemos que X não tem expectância finita e EX é indefinido.

Se X é uma variável aleatória não-negativa, geralmente indica-se por $EX < \infty$ o fato de que ela tem expectância finita.

Exemplo 2. Distribuição de Poisson. Suponha que X tem uma distribuição de Poisson de parâmetros λ. Então

$$EX = \sum_{j=1}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} = \sum_{j=1}^{\infty} \frac{\lambda^j}{(j-1)!} e^{-\lambda}$$
$$= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Exemplo 3. Distribuição geométrica. Suponha que X tem uma distribuição geométrica de parâmetro p. Obtenha EX.

Do exemplo (2) temos

$$EX = \sum_{j=0}^{\infty} jp(1-p)^j$$
$$= p(1-p) \sum_{j=0}^{\infty} j(1-p)^{j-1}$$
$$= -p(1-p) \sum_{j=0}^{\infty} \frac{d}{dp}(1-p)^j.$$

Como se pode diferenciar uma série geométrica, termo por termo, segue-se que

$$EX = -p(1-p)\frac{d}{dp}\sum_{j=0}^{\infty}(1-p)^j.$$

Usando a fórmula da soma de uma progressão geométrica, vemos que,

$$EX = -p(1-p)\frac{d}{dp}\left(\frac{1}{p}\right) = -p(1-p)\left(\frac{-1}{p^2}\right).$$

Conseqüentemente

$$EX = \frac{1-p}{p}.$$

Consideraremos a seguir um exemplo de uma densidade que não tem média finita.

Exemplo 4. Seja f a função definida em R por

$$f(x) = \begin{cases} \dfrac{1}{x(x+1)}, & x = 1, 2, \ldots, \\ 0, & \text{para outros valores de } x \end{cases}$$

Obviamente a função f satisfaz as propriedades (i) e (ii) da definição de funções de densidade dada no Capítulo 3. Para ver se f satisfaz a propriedade (iii), observamos que

$$\frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{x+1}$$

e portanto

$$\sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty}\left[\frac{1}{x} - \frac{1}{x+1}\right]$$

$$= (1 - 1/2) + (1/2 - 1/3) + \cdots = 1.$$

Assim f satisfaz (iii) e portanto é uma densidade. Mas f não tem média finita porque

$$\sum_{x=1}^{\infty} |x| f(x) = \sum_{x=1}^{\infty} \frac{1}{x+1}$$

e sabe-se muito bem que a série harmônica $\Sigma_{x=1}^{\infty} x^{-1}$ não converge.

4.2. PROPRIEDADES DA EXPECTÂNCIA

Freqüentemente desejamos determinar a expectância de uma variável aleatória como $Z = X_1 + X_2$ ou $Z = X^2$ que é em si mesma uma função $\varphi(\mathbf{X})$ do vetor aleatório $\mathbf{X}$. Naturalmente isto pode ser feito usando (2) se conhecemos a densidade f_Z de Z. Entretanto, com bastante freqüência a densidade de Z pode não ser conhecida ou, a determinação de EZ, a partir de uma dada densidade de Z, pode ser bastante difícil. O resultado que apresentaremos a seguir nos dá uma forma

de decidir se Z tem expectância finita, e se tiver, de determinar EZ diretamente em termos da densidade $f_{\mathbf{X}}$ e da função φ.

Antes de estabelecer este resultado, introduzimos uma convenção notacional. Seja $\mathbf{X}$ um vetor aleatório discreto r-dimensional que tem valores possíveis $\mathbf{x}_1, \mathbf{x}_2, \ldots$ e densidade f e seja φ uma função real qualquer definida em R^r. Então define-se $\Sigma_{\mathbf{x}}\ \varphi(\mathbf{x})f(\mathbf{x})$ como

(3) $$\sum_{\mathbf{x}} \varphi(\mathbf{x})f(\mathbf{x}) = \sum_j \varphi(\mathbf{x}_j)f(\mathbf{x}_j).$$

Teorema 1 Seja $\mathbf{X}$ um vetor aleatório discreto tendo densidade f e seja φ uma função real definida em R^r. Então a variável aleatória $Z = \varphi(\mathbf{X})$ tem expectância finita se e somente se

(4) $$\sum_{\mathbf{x}} |\varphi(\mathbf{x})|f(\mathbf{x}) < \infty$$

e quando (4) é verdadeiro,

(5) $$EZ = \sum_{\mathbf{x}} \varphi(\mathbf{x})f(\mathbf{x}).$$

Demonstração. Sejam $z_1, z_2, \ldots$ os distintos valores possíveis de Z e sejam $\mathbf{x}_1, \mathbf{x}_2, \ldots$ os distintos valores possíveis de $\mathbf{X}$. Para qualquer z_j existe pelo menos um $\mathbf{x}_i$ tal que, $z_j = \varphi(\mathbf{x}_i)$ mas, pode existir mais de um $\mathbf{x}_i$ que satisfaz essa condição. Seja A_j a coleção de tais x_i, isto é,

$$A_j = \{\mathbf{x}_i \mid \varphi(\mathbf{x}_i) = z_j\}.$$

Então $\{\mathbf{X} \in A_j\}$ e $\{Z = z_j\}$ representam exatamente os mesmos eventos. Assim

$$P(Z = z_j) = P(\mathbf{X} \in A_j) = \sum_{\mathbf{x} \in A_j} f_{\mathbf{X}}(\mathbf{x}).$$

Conseqüentemente,

$$\begin{aligned}\sum_j |z_j| f_Z(z_j) &= \sum_j |z_j| P(Z = z_j) \\ &= \sum_j |z_j| \sum_{\mathbf{x} \in A_j} f_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_j \sum_{\mathbf{x} \in A_j} |z_j| f_{\mathbf{X}}(\mathbf{x}).\end{aligned}$$

Como $\varphi(\mathbf{x}) = z_j$ para $\mathbf{x}$ em A_j, segue-se que

$$\sum_j |z_j| f_Z(z_j) = \sum_j \sum_{\mathbf{x} \in A_j} |\varphi(\mathbf{x})| f_{\mathbf{X}}(\mathbf{x}).$$

Por sua definição, os conjuntos A_j são disjuntos para valores distintos de j, e sua união é o conjunto de todos os valores possíveis de $\mathbf{X}$. Portanto

$$\sum_j |z_j| f_Z(z_j) = \sum_{\mathbf{x}} |\varphi(\mathbf{x})| f_{\mathbf{X}}(\mathbf{x}).$$

Isto mostra que Z tem expectância finita se, e somente se, (4) for verdadeiro.

Se Z possui expectância finita, então repetindo o mesmo argumento acima sem os sinais de módulo, concluímos que (5) é verdadeiro.

Seja X uma variável aleatória com densidade f e seja $\varphi(x) = |x|$. Então pelo Teorema 1, $|X|$ tem expectância finita se, e somente se, $\Sigma x\,|x|\,f(x) < \infty$. Mas, de acordo com nossa definição de expectância, X tem expectância se, e somente se, a mesma série converge. Vemos portanto que X tem expectância finita se, e somente se, $E|X| < \infty$.

Usaremos agora o Teorema 1 para estabelecer as importantes propriedades da expectância que seguem.

Teorema 2. Sejam X e Y duas variáveis aleatórias tendo expectância finita.

(i) Se c é uma constante e $P(X = c) = 1$, então $EX = c$.

(ii) Se c é uma constante, então cX tem expectância finita e $E(cX) = cEX$.

(iii) $X + Y$ tem expectância finita e $E(X + Y) = EX + EY$.

(iv) Suponha que $P(X \geqslant Y) = 1$. Então $EX \geqslant EY$; além do mais, $EX = EY$ se, e somente se, $P(X = Y) = 1$.

(v) $|EX| \leqslant E|X|$.

Demonstração. A demonstração de (i) é bastante simples. Se $P(X = c) = 1$, então X tem densidade $f_X(x) = 0$ para $x \neq c$ e $f_X(c) = 1$. Assim em virtude de (2)

$$EX = \sum_x x f_X(x) = c f_X(c) = c.$$

Para demonstrar (ii), seja $\varphi(x) = cx$ e observe que

$$\sum_x |cx| f_X(x) = |c| \sum_x |x| f_X(x) < \infty,$$

de modo que cX tem expectância finita. Assim de (5)

$$E(cX) = \sum_x (cx) f_X(x) = c \sum_x x f_X(x) = cEX.$$

Para estabelecer (iii) fazemos $\varphi(x, y) = x + y$ e seja f a densidade conjunta de X e Y. Então

$$\begin{aligned}\sum_{x,y} |x + y| f(x, y) &\leq \sum_{x,y} |x| f(x, y) + \sum_{x,y} |y| f(x, y) \\ &= \sum_x |x| \sum_y f(x, y) + \sum_y |y| \sum_x f(x, y) \\ &= \sum_x |x| f_X(x) + \sum_y |y| f_Y(y) < \infty\end{aligned}$$

e portanto $X + Y$ tem expectância finita. Aplicando (5) vemos que

$$\begin{aligned}E(X + Y) &= \sum_{x,y} (x + y) f(x, y) \\ &= \sum_{x,y} x f(x, y) + \sum_{x,y} y f(x, y) \\ &= EX + EY.\end{aligned}$$

Para demonstrar (iv) observe que $Z = X - Y = X + (-Y)$, e de acordo com (ii) e (iii) vemos que

$$EX - EY = E(X - Y) = EZ = \sum_z z f_Z(z).$$

Como $P(Z \geqslant 0) = P(X \geqslant Y) = 1$, os valores z_i que $Z = X - Y$ assume, devem ser todos não-negativos. Assim $\Sigma_z \, z f_Z(z) \geqslant 0$ e portanto $EX - EY \geqslant 0$. Isto conduz à primeira parte de (iv). Se $EX = EY$, então

$$0 = EZ = \sum_i z_i f_Z(z_i).$$

Mas a soma de termos não negativos pode ser zero somente se todos os termos individuais forem zero. Como $f_Z(z_i) > 0$, devemos ter $z_i = 0$. Assim o único valor possível de Z é 0 e conseqüentemente $P(Z = 0) = 1$.

Finalmente (v) decorre de (iv) e (ii) porque $-|X| \leqslant X \leqslant |X|$ e portanto $-E|X| \leqslant EX \leqslant E|X|$. Isto completa a demonstração do teorema.

Segue-se facilmente de (ii) e (iii) que se $X_1, \ldots, X_n$ são n variáveis aleatórias quaisquer que têm expectância finita e $c_1, \ldots, c_n$ são n constantes quaisquer, então

$$(6) \qquad E(c_1X_1 + \cdots + c_nX_n) = c_1EX_1 + \cdots + c_nEX_n.$$

É útil saber que uma variável aleatória limitada sempre tem expectância finita. Mais precisamente

Teorema 3. Seja X uma variável aleatória tal que $P(|X| \leqslant M) = 1$, para uma alguma constante M. Então X tem expectância finita e $|EX| \leqslant M$.

Demonstração. Sejam $x_1, x_2, \ldots$ os valores possíveis de X. Então $|x_i| \leqslant M$ para todo i. Realmente, se $|x_i| > M$ para algum valor possível de x_i, então

$$P(|X| > M) \geq P(|X| = |x_i|) > 0,$$

o que contradiz o fato de que $P(|X| \leqslant M) = 1$. Conseqüentemente

$$\sum_i |x_i| f(x_i) \leq M \sum_i f(x_i) \leq M,$$

de modo que X tem expectância finita. Além do mais, em virtude de (v) do Teorema 2.

$$|EX| \leq E|X| = \sum_i |x_i| f(x_i) \leq M.$$

Isto completa a demonstração.

Segue-se facilmente do Teorema 3 e de (iii) do Teorema 2 que, se X e Y são duas variáveis aleatórias tais que, Y tem expectância finita e $P(|X - Y| \leqslant M) = 1$, para alguma constante M, então X também tem expectância finita e $|EX - EY| \leqslant M$. Deixamos para o leitor a demonstração deste resultado.

Uma vez que a expectância da soma de duas variáveis aleatórias é a soma de suas expectâncias, podia-se supor que a expectância de um produto fosse o produto

das expectâncias. Pode-se ver que isto não é verdade, em geral, considerando a variável aleatória X que assume cada um dos valores 1 e -1 com probabilidade $1/2$ e fazendo $X = Y$. Então $EX = EY = 0$ mas $EXY = EX^2 = 1$.

Existe um caso importante em que esta regra do produto é válida. Trata-se do caso em que X e Y são variáveis aleatórias independentes. Estabelecemos formalmente este caso a seguir.

Teorema 4. Sejam X e Y duas variáveis aleatórias independentes tendo expectâncias finitas. Então XY tem expectância finita e

$$E(XY) = (EX)(EY). \tag{7}$$

Demonstração. Observe que a densidade conjunta de X e Y é $f_X(x)f_Y(y)$. uma vez que X e Y são independentes. Assim

$$\sum_{x,y} |xy| f(x, y) = \sum_{x,y} |x|\,|y| f_X(x) f_Y(y)$$

$$= \left(\sum_x |x| f_X(x)\right)\left(\sum_y |y| f_Y(y)\right) < \infty,$$

de modo que XY tem expectância finita. Usando o Teorema 1 concluímos que

$$E(XY) = \sum_{x,y} (xy) f_X(x) f_Y(y)$$

$$= \left[\sum_x x f_X(x)\right]\left[\sum_y y f_Y(y)\right] = (EX)(EY).$$

A recíproca desta propriedade não é verdadeira; duas variáveis aleatórias podem ser tais que $E(XY) = (EX)(EY)$ mesmo que X e Y não sejam independentes.

Exemplo 5. Suponha que (X, Y) assume os valores $(1,0)$, $(0,1)$, $(-1,0)$ e $(0,-1)$ com probabilidades iguais. Então $EX = EY = 0$. Como $XY = 0$, segue-se que $E(XY) = 0$ e portanto $E(XY) = (EX)(EY)$. Para ver que X e Y não são independentes, observe por exemplo que $P(X = 0) = P(Y = 0) = 1/2$ enquanto $P(X = 0, Y = 0) = 0$. Assim

$$P(X = 0, Y = 0) \neq P(X = 0)P(Y = 0).$$

Freqüentemente é mais fácil determinar expectâncias usando as propriedades dadas no Teorema 2 do que usando diretamente a definição. Ilustraremos agora essa técnica com diversos exemplos.

Exemplo 6. Distribuição binomial. Já sabemos do Exemplo 1 que a média da distribuição binomial de parâmetros n e p é np. Podemos também obter este resultado de uma maneira muito simples, usando a propriedade de que a expectância de uma soma é a soma das expectâncias ((iii) do Teorema 2). Para isto, sejam $X_1, \ldots, X_n$ n variáveis aleatórias independentes de Bernoulli de parâmetro p

e seja $S_n = X_1 + \cdots + X_n$. Então S_n tem a distribuição binomial de parâmetros n e p. Da primeira parte do Exemplo 1

$$EX_i = p, 1 \leq i \leq n, \quad \text{e portanto}$$

$$ES_n = E(X_1 + \cdots + X_n) = \sum_{i=1}^{n} EX_i = np.$$

Exemplo 7. Distribuição hipergeométrica. Suponha que temos uma população de r objetos dos quais r_1 são do tipo um e $r - r_1$, são do tipo dois. Extrai-se desta população uma amostra sem reposição de tamanho n. Seja S_n o número de objetos do tipo um na amostra. Determine ES_n.

Sabemos que S_n tem distribuição hipergeométrica, de modo que poderíamos determinar ES_n usando (2). Entretanto, é muito mais simples proceder introduzindo as variáveis aleatórias indicadoras $X_1, \ldots, X_n$ da seguinte forma. A variável aleatória $X_i = 1$ se, e somente se, o i-ésimo elemento da amostra é do tipo um. Então

$$ES_n = \sum_{i=1}^{n} EX_i = n\frac{r_1}{r}.$$

Mas $S_n = X_1 + \cdots + X_n$, de modo que usando (iii) do Teorema 2 vemos que

$$EX_i = P(X_i = 1) = \frac{r_1}{r}.$$

Observe que as variáveis aleatórias X_i, $1 \leqslant i \leqslant n$, *não* são independentes.

Exemplo 8. Suponha que temos uma população de r objetos distintos numerados de 1 a r. Extrai-se os objetos com reposição até obter exatamente $k \leqslant r$ objetos distintos. Seja S_k o tamanho da amostra necessária. Determine ES_k.

É claro que $S_1 = 1$ e portanto $ES_1 = 1$. Suponha que $k \geqslant 2$ e seja $X_i = S_{i+1} - S_i, i = 1, 2, \ldots, k-1$. Então claramente $S_k = 1 + X_1 + \cdots + X_{k-1}$. Mas X_i é o número de objetos que se deve extrair até que o $(i + 1)$-ésimo objeto novo entre na amostra, após o i-ésimo objeto novo ter entrado na amostra. Um pouco de raciocínio mostra que o evento $\{X_i = n\}$ ocorre se, e somente se, os primeiros $n - 1$ elementos, extraídos depois que o i-ésimo objeto novo entra na amostra, repetem um dos i objetos anteriores, e o n-ésimo elemento, extraído depois do i-ésimo objeto novo, é diferente dos i objetos anteriores. Assim, como as extrações são independentes,

$$P(X_i = n) = \left(\frac{i}{r}\right)^{n-1}\left(1 - \frac{i}{r}\right), \qquad n = 1, 2, \ldots.$$

Isto mostra que a variável aleatória $X_i - 1$ distribui-se geometricamente com parâmetro $p_i = 1 - (i/r)$. Portanto, de acordo com o Exemplo 3, $E(X_i - 1) = p_i^{-1}(1 - p_i)$, e

$$EX_i = p_i^{-1}(1 - p_i) + 1 = p_i^{-1} = (1 - i/r)^{-1} = r(r - i)^{-1}.$$

Conseqüentemente,

$$(8) \qquad ES_k = 1 + \sum_{i=1}^{k-1} \left(\frac{r}{r-i}\right)$$

$$= \sum_{i=0}^{k-1} \left(\frac{r}{r-i}\right)$$

$$= r\left(\frac{1}{r} + \frac{1}{r-1} + \cdots + \frac{1}{r-k+1}\right).$$

Registramos para uso futuro, que se torna claro da construção de X_i, que elas são variáveis aleatórias mutuamente independentes.

Vimos no capítulo anterior que variáveis aleatórias inteiras não-negativas X desempenham um papel proeminente. Para estas variáveis aleatórias pode-se freqüentemente aplicar o teorema seguinte para decidir se X tem expectância finita e determinar a expectância de X.

Teorema 5. Seja X uma variável aleatória inteira não-negativa. Então X tem expectância finita se, e somente se, a série $\Sigma_{x=1}^{\infty} P(X \geqslant x)$ converge. Se a série converge, então

$$(9) \qquad EX = \sum_{x=1}^{\infty} P(X \geq x).$$

Demonstração. Mostraremos que

$$(10) \qquad \sum_{x=1}^{\infty} xP(X = x) = \sum_{x=1}^{\infty} P(X \geq x),$$

do qual segue-se imediatamente o teorema. Para isso escrevemos inicialmente o primeiro membro de (10) como

$$\sum_{x=1}^{\infty} P(X = x) \sum_{y=1}^{x} 1.$$

É possível inverter a ordem dos somatórios e reescrever esta expressão como

$$\sum_{y=1}^{\infty} \sum_{x=y}^{\infty} P(X = x) = \sum_{y=1}^{\infty} P(X \geq y).$$

Substituindo a variável auxiliar y pela variável auxiliar x no segundo membro desta igualdade, obtemos o segundo membro de (10). Isto mostra que (10) é verdadeiro como desejado.

Para uma aplicação elementar deste teorema, suponha que X é uma variável aleatória que se distribui geometricamente com parâmetro p. Então $P(X \geqslant x) = (1-p)^x$ e de acordo com o teorema

$$EX = \sum_{x=1}^{\infty} (1-p)^x = (1-p) + (1-p)^2 = \cdots = p^{-1}(1-p).$$

Isto está de acordo com o resultado obtido no Exemplo 3.

4.3. MOMENTOS

Seja X uma variável aleatória discreta e seja $r \geqslant 0$ um número inteiro. Dizemos que X tem um momento de ordem r se X^r tem expectância finita. Neste caso definimos o r-ésimo momento de X como EX^r. Se X tem um momento de ordem r, então o r-ésimo momento de $X - \mu$, onde μ é a média de X, chama-se r-ésimo *momento central* (ou r-ésimo momento em torno da média) de X. Podemos determinar pelo Teorema 1 o r-ésimo momento e o r-ésimo momento central de X diretamente da densidade f pelas fórmulas

(11) $$EX^r = \sum_x x^r f(x)$$

e

(12) $$E(X - \mu)^r = \sum_x (x - \mu)^r f(x).$$

De acordo com (11) e (12), o r-ésimo momento e o r-ésimo momento central são determinados pela densidade f, de modo que faz sentido falar neles como o r-ésimo momento e o r-ésimo momento central desta densidade.

Suponha que X tem um momento de ordem r, então X tem um momento de ordem k para todo $k \leqslant r$. Para ver isto, observe que se $|x| \leqslant 1$, então

$$|x^k| = |x|^k \leq 1$$

enquanto para $|x| > 1$,

$$|x|^k \leq |x|^r.$$

Assim, em ambos os casos é sempre verdade que

$$|x|^k \leq |x|^r + 1.$$

Então, pelo teorema da comparação para convergência de séries, vemos que

$$\sum_x |x|^k f(x) \leq \sum_x [|x|^r + 1] f(x) = E(|X|^r) + 1 < \infty,$$

de modo que X^k tem expectância finita.

Por outro lado, como mostramos no Exemplo 4, uma variável aleatória X pode não ter nem mesmo o primeiro momento. Uma simples modificação deste exemplo mostra que uma variável aleatória pode ter um momento de ordem r mas não ter momentos de ordens superiores. (Ver Exercício 9.)

O primeiro momento ($r = 1$) é simplesmente a média de X. Em geral, quanto maior o número de momentos de X que conhecemos, maior é a informação que temos sobre a distribuição de X; entretanto, os dois primeiros momentos são de maior interesse nas aplicações.

Sabemos pela propriedade (iii) do Teorema 2 que se X e Y têm ambos um primeiro momento finito, então $X + Y$ também o tem. Mostraremos a seguir que esta desejável propriedade mantém-se válida para momentos de ordem r.

Teorema 6. Se as variáveis aleatórias X e Y têm momentos de ordem r, então $X + Y$ também tem momento de ordem r.

Demonstração. Este teorema baseia-se na simples desigualdade a seguir. Para qualquer número inteiro não-negativo $j \leqslant r$,

$$|x|^j|y|^{r-j} \leq |x|^r + |y|^r, \qquad x, y \in R. \tag{13}$$

Para ver isto, observe que se $|x| \leqslant |y|$, então $|x|^j|y|^{r-j} \leqslant |y|^j|y|^{r-j} = |y|^r \leqslant |x|^r + |y|^r$; enquanto se $|x| \geqslant |y|$, então $|x|^j|y|^{r-j} \leqslant |x|^r \leqslant |x|^r + |y|^r$. Assim (13) é verdadeiro. Usando (13) e o teorema binomial, vemos que

$$\begin{aligned} |x + y|^r &\leq (|x| + |y|)^r \\ &= \sum_{j=0}^{r} \binom{r}{j} |x|^j|y|^{r-j} \\ &\leq \sum_{j=0}^{r} \binom{r}{j} (|x|^r + |y|^r). \end{aligned}$$

Mas

$$\sum_{j=0}^{r} \binom{r}{j} = 2^r$$

porque

$$2^r = (1 + 1)^r = \sum_{j=0}^{r} \binom{r}{j} 1^j 1^{r-j} = \sum_{j=0}^{r} \binom{r}{j}.$$

Conseqüentemente

$$|x + y|^r \leq 2^r(|x|^r + |y|^r).$$

Seja f a densidade conjunta de X e Y. Então

$$\begin{aligned} \sum_{x,y} |x + y|^r f(x, y) &\leq 2^r \sum_{x,y} (|x|^r + |y|^r) f(x, y) \\ &= 2^r E(|X|^r + |Y|^r) \\ &= 2^r(E|X|^r + E|Y|^r) < \infty. \end{aligned}$$

Portanto, pelo Teorema 1, $(X + Y)^r$ tem expectância finita.

Segue-se facilmente por indução que se $X_1, X_2, \ldots, X_n$ tem todos um momento de ordem r, então $X_1 + \cdots + X_n$ também o tem.

Seja X uma variável aleatória que tem um momento finito de segunda ordem. Então a *variância de* X, representada por $\operatorname{Var} X$ ou $V(X)$, é definida por

$$\operatorname{Var} X = E[(X - EX)^2].$$

Expandindo o segundo membro vemos que

$$\begin{aligned} \operatorname{Var} X &= E[X^2 - (2X)(EX) + (EX)^2] \\ &= EX^2 - 2(EX)^2 + (EX)^2. \end{aligned}$$

Em outras palavras

(14) $$\operatorname{Var} X = EX^2 - (EX)^2.$$

Representa-se freqüentemente EX por μ e Var X por σ^2. O número não-negativo $\sigma = \sqrt{\operatorname{Var} X}$ chama-se *desvio padrão* de X ou de f_X.

De acordo com a nossa discussão anterior, μ é o valor médio da variável aleatória X. Um dos usos da variância é como uma medida da dispersão da distribuição de X em torno da média. Quanto maior a tendência de X de se afastar do seu valor médio, maior a tendência de $(X - \mu)^2$ de assumir valor grande, e portanto, maior será a variância.

Por outro lado, Var $X = 0$ se, e somente se, X é uma constante. Para ver isto, observe que se $P(X = c) = 1$ para alguma constante c, então $EX = c$ e Var $X = 0$. Reciprocamente, se Var $X = 0$, então $P((X - EX)^2 = 0) = 1$, portanto, $E[(X - EX)^2] = 0$. Conseqüentemente $P(X = EX) = 1$.

O problema seguinte, que é de interesse em estatística, ilustra um uso alternativo da média e variância. Seja X uma variável aleatória que tem um segundo momento finito, e suponha que desejamos determinar o valor de a que minimiza $E(X - a)^2$. Tal valor nos daria uma constante que melhor se ajusta a X se o erro fosse medido pelo desvio quadrático médio.

Uma forma de resolver este problema é pelo uso de cálculo. Observe que

$$E(X - a)^2 = EX^2 - 2aEX + a^2.$$

Se diferenciarmos em relação a a e igualarmos a derivada a zero, veremos que $a = EX$. Como a segunda derivada é positiva (na realidade, ela é igual a 2), o ponto corresponde a um mínimo, e o valor mínimo é Var X.

Existe uma segunda forma de resolver este problema que é também importante compreender. Observe que

$$\begin{aligned}(X - a)^2 &= [(X - \mu) + (\mu - a)]^2 \\ &= (X - \mu)^2 + 2(X - \mu)(\mu - a) + (\mu - a)^2.\end{aligned}$$

Como $E(X - \mu) = 0$, o termo de produto cruzado tem expectância nula e portanto,

(15) $$\begin{aligned}E(X - a)^2 &= E(X - \mu)^2 + (\mu - a)^2 \\ &= \operatorname{Var} X + (\mu - a)^2.\end{aligned}$$

Torna-se claro que de (15) que $E(X - a)^2$ alcança um mínimo quando $\mu = a$, e que o valor deste mínimo é Var X.

Freqüentemente podemos determinar mais simplesmente os momentos de uma variável aleatória inteira não negativa pela diferenciação de sua função geratriz de probabilidade Φ_X. Suponha por simplicidade que

$$\sum_{x=0}^{\infty} f_X(x) t_0^x < \infty$$

para algum $t_0 > 1$. Então podemos considerar Φ_X bem definido em $-t_0 < t < t_0$ através de

$$\Phi_X(t) = \sum_{x=0}^{\infty} f_X(x)t^x, \qquad -t_0 < t < t_0.$$

Podemos diferenciar $\Phi_X(t)$ qualquer número de vezes, diferenciando, termo a termo, a correspondente série de potências. Em particular

$$\Phi'_X(t) = \sum_{x=1}^{\infty} x f_X(x)t^{x-1}, \qquad -t_0 < t < t_0,$$

e

$$\Phi''_X(t) = \sum_{x=2}^{\infty} x(x-1) f_X(x)t^{x-2}, \qquad -t_0 < t < t_0.$$

Em virtude das hipóteses sobre t_0, podemos fazer $t = 1$ nestas fórmulas, obtendo

$$\Phi'_X(1) = \sum_{x=1}^{\infty} x f_X(x) = EX$$

e

$$\Phi''_X(1) = \sum_{x=2}^{\infty} x(x-1) f_X(x) = EX(X-1).$$

Assim, pode-se obter de Φ_X a média e a variância de X por meio das fórmulas

$$EX = \Phi'_X(1)$$

e

$$\text{Var}\, X = EX^2 - (EX)^2 = \Phi''_X(1) + \Phi'_X(1) - (\Phi'_X(1))^2.$$

Pode-se desenvolver fórmulas semelhantes para os outros momentos de X em termos de derivadas superiores de Φ_X no ponto $t = 1$.

Ilustraremos a seguir o uso dessas fórmulas.

Exemplo 9. Distribuição binomial negativa. Seja X uma variável aleatória com distribuição binomial negativa de parâmetros α e p. Obtenha a média e a variância de X.

De acordo com o Exemplo 17 do Capítulo 3 sabemos que a função geratriz de X é dada por $\Phi_X(t) = p^\alpha [1 - t(1-p)]^{-\alpha}$. Conseqüentemente

$$\Phi'_X(t) = \alpha p^\alpha [1 - t(1-p)]^{-(\alpha+1)}(1-p)$$

e

$$\Phi''_X(t) = (\alpha+1)\alpha p^\alpha [1 - t(1-p)]^{-(\alpha+2)}(1-p)^2.$$

Assim

$$\Phi'_X(1) = \alpha \left(\frac{1-p}{p}\right)$$

e

$$\Phi''_X(1) = (\alpha+1)\alpha \left(\frac{1-p}{p}\right)^2.$$

Portanto, $EX = \alpha p^{-1}(1-p)$ e

$$\begin{aligned}\text{Var } X &= (\alpha+1)\alpha\left(\frac{1-p}{p}\right)^2 + \alpha\left(\frac{1-p}{p}\right) - \alpha^2\left(\frac{1-p}{p}\right)^2\\ &= \alpha\frac{1-p}{p^2}.\end{aligned}$$

Em particular, se X tem uma distribuição geométrica de parâmetro p, então $EX = p^{-1}(1-p)$ (como já vimos) e $\text{Var } X = p^{-2}(1-p)$.

Exemplo 10. Distribuição de Poisson. Suponha que X tem uma distribuição de Poisson de parâmetro λ. Obtenha a média e a variância de X.

No Exemplo 18 do Capítulo 3 vimos que $\Phi_X(t) = e^{\lambda(t-1)}$. Assim

$$\Phi_X'(t) = \lambda e^{\lambda(t-1)}$$

e

$$\Phi_X''(t) = \lambda^2 e^{\lambda(t-1)}.$$

Conseqüentemente $\Phi_X'(1) = \lambda$ e $\Phi_X''(1) = \lambda^2$. Segue-se imediatamente que

$$EX = \lambda,$$

concorda com o resultado obtido no Exemplo 2, e

$$\text{Var } X = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Isto mostra que se X tem uma distribuição de Poisson de parâmetro λ, então a média e a variância de X são iguais a λ.

4.4. VARIÂNCIA DE UMA SOMA

Sejam X e Y duas variáveis aleatórias que têm segundos momentos finitos. Então $X + Y$ tem segundo momento finito e portanto variância finita. Temos que

$$\begin{aligned}\text{Var }(X+Y) &= E[(X+Y) - E(X+Y)]^2\\ &= E[(X-EX) + (Y-EY)]^2\\ &= E(X-EX)^2 + E(Y-EY)^2\\ &\quad + 2E[(X-EX)(Y-EY)]\\ &= \text{Var } X + \text{Var } Y + 2E[(X-EX)(Y-EY)].\end{aligned}$$

Assim, ao contrário da média, a variância da soma de duas variáveis aleatórias não é em geral igual à soma das variâncias.

A quantidade

$$E[(X-EX)(Y-EY)]$$

chama-se *covariância* de X e Y e representa-se por $\text{Cov}(X, Y)$.

Temos assim a importante fórmula

$$\text{(16)} \qquad \text{Var }(X+Y) = \text{Var } X + \text{Var } Y + 2\,\text{Cov}(X, Y).$$

Uma vez que

$$(X - EX)(Y - EY) = XY - (Y)(EX) - X(EY) + (EX)(EY),$$

tomando as expectâncias, vemos que

$$\text{(17)} \qquad \text{Cov}\,(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - (EX)(EY).$$

Desta forma, torna-se claro que $\text{Cov}\,(X, Y) = 0$ sempre que X e Y são independentes. (O Exemplo 5 mostra que a recíproca é falsa.) Vemos de (16) que se X e Y são variáveis aleatórias independentes, tendo segundos momentos finitos, então $\text{Var}\,(X + Y) = \text{Var}\,X + \text{Var}\,Y$.

Em particular se $P(Y = c) = 1$ para alguma constante c, então X e Y são independentes e a variância de Y é zero; conseqüentemente

$$\text{(18)} \qquad \text{Var}\,(X + c) = \text{Var}\,X + \text{Var}\,(c) = \text{Var}\,X.$$

De uma forma mais geral, se $X_1, X_2, \ldots, X_n$ são n variáveis aleatórias, cada uma com segundo momento finito, então

$$\text{(19)} \qquad \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}\,X_i + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \text{Cov}\,(X_i, X_j),$$

e em particular, se $X_1, \ldots, X_n$ são mutuamente independentes, então

$$\text{(20)} \qquad \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}\,X_i.$$

Pode-se obter estas fórmulas através de um desenvolvimento direto parecido com (porém mais complicado que) o usado no caso $n = 2$, ou pode-se estabelecê-las por indução sobre n, partindo de $n = 2$.

Em particular, se $X_1, \ldots, X_n$ são variáveis aleatórias independentes tendo variância comum σ^2 (se elas têm a mesma densidade, por exemplo), então

$$\text{(21)} \qquad \text{Var}\,(X_1 + \cdots + X_n) = n\,\text{Var}\,X_1 = n\sigma^2.$$

Outro fato elementar, porém, bastante útil é que $\text{Var}\,(aX) = a^2\,\text{Var}\,X$. Deixamos para o leitor a verificação deste fato.

Exemplo 11. Distribuição binomial. Sejam, $X_1, \ldots, X_n$, n variáveis aleatórias independentes de Bernoulli tendo cada uma a mesma probabilidade p de assumir o valor 1. Então (ver Exemplo 6) a soma $S_n = X_1 + \cdots + X_n$ distribui-se binomialmente com parâmetros n e p. Mostramos anteriormente que $ES_n = np$. Usando (21) obtemos de imediato que

$$\text{Var}\,S_n = n\,\text{Var}\,X_1.$$

Mas $X_1^2 = X_1$ porque X_1 é 0 ou 1. Assim $EX_1^2 = EX_1 = p$ e portanto

$$\text{Var}\,X_1 = EX_1^2 - (EX_1)^2 = p - p^2 = p(1 - p).$$

Conseqüentemente $\text{Var}\,S_n = np(1 - p)$.

Em resumo então, a média de uma variável binomialmente distribuída é np e sua variância é $np(1-p)$.

Exemplo 12. Distribuição hipergeométrica. Considere a situação do Exemplo 7. Desejamos agora determinar Var S_n para obter a variância de uma distribuição hipergeométrica. Para isso usaremos a equação (19).

Para os indicadores independentes $X_1, \ldots, X_n$, obtivemos anteriormente que

$$P(X_i = 1) = EX_i = \frac{r_1}{r}.$$

Como $X_1^2 = X_i$ vemos que

$$\text{Var } X_i = EX_i^2 - (EX_i)^2 = \left(\frac{r_1}{r}\right) - \left(\frac{r_1}{r}\right)^2$$

$$= \left(\frac{r_1}{r}\right)\left(1 - \frac{r_1}{r}\right).$$

A seguir determinamos as covariâncias. Suponha que $1 \leqslant i < j \leqslant n$. Como $X_iX_j = 0$ a menos que X_i e X_j. Sejam ambos iguais a 1, temos que

$$EX_iX_j = P(X_i = 1, X_j = 1) = \left(\frac{r_1}{r}\right)\left(\frac{r_1 - 1}{r - 1}\right).$$

Assim

$$\text{Cov } (X_i, X_j) = E(X_iX_j) - (EX_i)(EX_j)$$

$$= \frac{r_1(r_1 - 1)}{r(r - 1)} - \left(\frac{r_1}{r}\right)^2$$

$$= \left(\frac{r_1}{r}\right)\left(\frac{r_1 - 1}{r - 1} - \frac{r_1}{r}\right)$$

$$= \left(\frac{r_1}{r}\right)\frac{r_1 - r}{r(r - 1)},$$

e portanto

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{Cov } (X_i, X_j) = \frac{n(n - 1)}{2}\left(\frac{r_1}{r}\right)\frac{r_1 - r}{r(r - 1)}.$$

Segue-se então de (19) que

$$\text{Var } S_n = n\,\frac{r_1(r - r_1)}{r^2} - n(n - 1)\frac{r_1(r - r_1)}{r^2(r - 1)}$$

$$= n\left(\frac{r_1}{r}\right)\left(1 - \frac{r_1}{r}\right)\left(1 - \frac{n - 1}{r - 1}\right).$$

É interessante comparar a média e variâncias da distribuição hipergeométrica com as da distribuição binomial tendo a mesma probabilidade de sucesso $p = (r_1/r)$. Suponha que temos uma população de r objetos dos quais r_1 são do tipo um e

$r - r_1$ são do tipo dois. Extrai-se uma amostra de tamanho n da população. Seja Y o número de objetos do tipo um na amostra.

Se a amostragem é *com reposição,* Y distribui-se binomialmente com parâmetros n e $p = (r_1/r)$, e portanto

$$EY = n\left(\frac{r_1}{r}\right) \quad \text{e} \quad \text{Var } Y = n\left(\frac{r_1}{r}\right)\left(1 - \frac{r_1}{r}\right).$$

Por outro lado, se a amostragem é *sem reposição,* Y tem a distribuição hipergeométrica, e

$$EY = n\left(\frac{r_1}{r}\right) \quad \text{e} \quad \text{Var } Y = n\left(\frac{r_1}{r}\right)\left(1 - \frac{r_1}{r}\right)\left(1 - \frac{n-1}{r-1}\right).$$

A média é a mesma nos dois casos, mas a variância é menor na amostragem sem reposição. Intuitivamente, quanto maior a proximidade entre n e r, mais determinístico se torna Y, quando a amostragem é feita sem reposição. Realmente, se $n = r$, a variância é zero e $P(Y = r_1) = 1$. Mas se r é grande em comparação a n, de forma que (n/r) se aproxima de zero, a razão das variâncias obtidas nas amostragens com e sem reposição é próxima de um. Isto era de se esperar, pois, para n fixo e r grande, existe pouca diferença entre amostragem com reposição e amostragem sem reposição.

4.5. COEFICIENTE DE CORRELAÇÃO

Sejam X e Y duas variáveis aleatórias tendo variâncias finitas e não-nulas. Uma medida do grau de dependência entre duas variáveis aleatórias é o coeficiente de correlação $\rho(X, Y)$ definida por

$$\rho = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{(\text{Var } X)(\text{Var } Y)}}. \tag{22}$$

Diz-se que estas variáveis aleatórias não são correlacionadas se $\rho = 0$. Como $\text{Cov}(X, Y) = 0$ se X e Y são independentes, vemos de imediato que variáveis aleatórias independentes não são correlacionadas. É também possível que variáveis aleatórias dependentes não sejam correlacionadas, como se pode ver do Exemplo 5.

É importante para aplicações em estatística saber que o coeficiente de correlação está sempre entre -1 e 1, e que $|\rho| = 1$ se, e somente se, $P(X = aY) = 1$ para alguma constante a. Estes fatos são conseqüências imediatas da desigualdade básica a seguir chamada *desigualdade de Schwarz.*

Teorema 7. A desigualdade de Schwarz. Suponha que X e Y têm segundos momentos finitos. Então

$$[E(XY)]^2 \leq (EX^2)(EY^2). \tag{23}$$

Além do mais a igualdade prevalece em (23) se e somente se $P(Y = 0) = 1$ ou $P(X = aY) = 1$ para alguma constante a.

Demonstração. Se $P(Y = 0) = 1$, então $P(XY = 0) = 1$, $EXY = 0$, e $EY^2 = 0$; assim, neste caso (23) se verifica com a igualdade. Também se $P(X = aY) = 1$, um simples desenvolvimento mostrará que ambos os membros de (23) são iguais a $(a^2 EY^2)^2$.

Mostraremos agora que (23) é sempre verdadeiro. Em virtude da discussão acima podemos supor que $P(Y = 0) < 1$ e portanto $EY^2 > 0$. A demonstração baseia-se em um artifício simples porém engenhoso. Observe que para qualquer número real λ

$$0 \leq E(X - \lambda Y)^2 = \lambda^2 EY^2 - 2\lambda EXY + EX^2.$$

Trata-se de uma função quadrática de λ. Como o coeficiente EY^2 de λ^2 é positivo, obtém-se o mínimo para algum valor de λ, digamos $\lambda = a$, que pode ser obtido pelo método usual de cálculo, ou seja, igualando a primeira derivada a zero e resolvendo a equação resultante. A resposta é $a = [E(XY)][EY^2]^{-1}$. Como o valor resultante da função é

$$(24) \qquad 0 \leq E(X - aY)^2 = EX^2 - \frac{[E(XY)]^2}{EY^2}$$

segue-se que (23) é válido. Se a igualdade prevalecer na desigualdade de Schwarz, então vemos de (24) que $E(X - aY)^2 = 0$, de modo que

$$P[(X - aY) = 0] = 1.$$

Isto completa a demonstração.

Aplicando a desigualdade de Schwarz às variáveis aleatórias $(X - EX)$ e $(Y - EY)$ vemos que

$$(E[(X - EX)(Y - EY)])^2 \leq [E(X - EX)^2][E(Y - EY)^2];$$

isto é,

$$[\text{Cov}\,(X, Y)]^2 \leq (\text{Var}\,X)(\text{Var}\,Y).$$

Assim, pela definição de ρ

$$|\rho(X, Y)| \leq 1.$$

Vemos também pelo Teorema 7 que $|\rho| = 1$ se, e somente se, $P(X = aY) = 1$ para alguma constante a.

O coeficiente de correlação é de uso limitado na teoria da probabilidade. Ele ocorre principalmente em estatística, e sua discussão mais detalhada será adiada para o Volume II.

4.6. DESIGUALDADE DE CHEBYSHEV

Seja X uma variável aleatória não negativa tendo expectância finita e seja t um número real positivo. Defina a variável aleatória Y fazendo $Y = 0$ se $X < t$ e $Y = t$ se $X \geqslant t$. Então Y é uma variável aleatória discreta tendo dois valores

possíveis 0 e t que ela assume com probabilidades $P(Y = 0) = P(X < t)$ e $P(Y = t) = P(X \geqslant t)$ respectivamente.

Assim

$$EY = tP(Y = t) + 0 \cdot P(Y = 0) = tP(Y = t) = tP(X \geq t).$$

Mas claramente $X \geqslant Y$ e portanto $EX \geqslant EY$. Assim

$$EX \geq EY = tP(X \geq t)$$

ou

$$P(X \geq t) \leq \frac{EX}{t}. \tag{25}$$

Pode-se deduzir de (25) uma considerável variedade de desigualdades úteis. A mais importante delas é a desigualdade de Chebyshev.

Desigualdade de Chebyshev. Seja X uma variável aleatória com média μ e variância finita σ^2. Então, para qualquer número real $t > 0$

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}. \tag{26}$$

Para demonstrar (26), aplicamos (25) à variável aleatória não-negativa $(X - \mu)^2$ e o número t^2. Concluímos que

$$P((X - \mu)^2 \geq t^2) \leq \frac{E(X - \mu)^2}{t^2} = \frac{\sigma^2}{t^2}.$$

Como $(X - \mu)^2 \geqslant t^2$ se, e somente se, $|X - \mu| \geqslant t$ vemos que (26) é verdadeiro.

A desigualdade de Chebyshev estabelece um limite superior em termos de Var X e t para a probabilidade de que X se desvie de sua média por mais t unidades. Sua virtude reside em sua grande generalidade. Não se faz nenhuma restrição sobre a distribuição de X, exceto que ela possua variância finita. Esta desigualdade é o ponto de partida para vários desenvolvimentos teóricos. Para a maioria das distribuições que ocorrem na prática, existem limites muito mais rigorosos para $P(|X - \mu| \geqslant t)$ do que o dado pela desigualdade de Chebyshev, entretanto, exemplos mostram que em geral o limite estabelecido pela desigualdade de Chebyshev não pode ser melhorado (ver Exercício 26).

Sejam $X_1, \ldots, X_n$ n variáveis aleatórias independentes que têm a mesma distribuição, pode-se pensar nestas variáveis aleatórias como n medidas independentes de alguma grandeza, que se distribui, de acordo com sua distribuição comum. Neste sentido falamos às vezes nas variáveis aleatórias $X_1, \ldots, X_n$ como, constituindo uma amostra aleatória de tamanho n desta distribuição.

Suponha que a distribuição comum dessas variáveis aleatórias tenha média finita μ. Então, para n suficientemente grande, esperaríamos que sua média aritmética $S_n/n = (X_1 + \cdots + X_n)/n$ se aproxime de μ. Se X_i tiver também variância finita, então

$$\text{Var}\left(\frac{S_n}{n}\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

e assim, Var $(S_n/n) \to 0$ quando $n \to \infty$. Como foi discutido na Seção 4.3., isto implica que, à medida que n cresce, a distribuição de (S_n/n) se concentra mais em torno de sua média μ. De modo mais preciso, aplicando a desigualdade de Chebyshev obtemos a desigualdade

$$(27) \qquad P\left(\left|\frac{S_n}{n} - \mu\right| \geq \delta\right) \leq \frac{\text{Var}(S_n/n)}{\delta^2} = \frac{\sigma^2}{n\delta^2}.$$

Em particular, segue-se de (27) que para qualquer $\delta > 0$

$$(28) \qquad \lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \delta\right) = 0.$$

Podemos interpretar (28) da seguinte forma. Pode-se pensar no número δ como a precisão desejada na aproximação de μ por S_n/n. A Equação (28) garante-nos que, por menor que seja o valor escolhido para δ, a probabilidade de que S_n/n aproxime μ com esta precisão, isto é, $P(|(S_n/n) - \mu| < \delta)$, converge para 1 à medida que o número de observações se torna grande. Este resultado chama-se Lei Fraca dos Grandes Números. Nós demonstramos a lei fraca apenas sob a hipótese de que a variância de X_i é finita. Na realidade isto não é necessário; tudo que é necessário é que X_i tenha média finita. Enunciamos este resultado mais geral no teorema a seguir. A demonstração será dada no Capítulo 8.

Teorema 8 Lei Fraca dos Grandes Números. Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes tendo uma distribuição comum com média finita μ e seja $S_n = X_1 + \cdots + X_n$. Então para qualquer $\delta > 0$

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq \delta\right) = 0.$$

A lei fraca é válida sempre que X_i tem média finita. Entretanto, se elas têm também variância finita, então (27) se verifica. Esta é uma asserção mais precisa uma vez que ela nos dá um limite superior para $P\left(\left|\frac{S_n}{n} - \mu\right| \geq \delta\right)$ em termos de n. Ilustraremos agora o uso de (27) aplicando-a a variáveis aleatórias com distribuição binomial.

Suponha que $X_1, \ldots, X_n$ sejam n variáveis aleatórias independentes de Bernoulli assumindo o valor 1 com probabilidade comum p. Então $\mu = p$ e $\sigma^2 = p(1-p)$. Assim, (27) mostra que

$$(29) \qquad P\left(\left|\frac{S_n}{n} - p\right| \geq \delta\right) \leq \frac{p(1-p)}{n\delta^2}.$$

Como $p(1-p) \leqslant 1/4$ se $0 < p < 1$ (porque métodos habituais de cálculo mostram que $p(1-p)$ alcança o seu máximo para $p = 1/2$), segue-se que, independentemente do valor de p,

$$(30) \qquad P\left(\left|\frac{S_n}{n} - p\right| \geq \delta\right) \leq \frac{1}{4n\delta^2}.$$

A Equação (29) é útil quando conhecemos o valor de p, enquanto (30) nos dá um limite para $P\left(\left|\frac{S_n}{n} - p\right| \geq \delta\right)$, que é válida para qualquer valor de p. Se p estiver próximo de 1/2, (29) e (30) não diferem muito, mas se p estiver longe de 1/2, a estimativa dada por (29) pode ser muito melhor. (Na realidade, mesmo os limites dados por (29) são bastante pobres. No Capítulo 7 discutiremos outro método que produz estimativas muito melhores).

Suponha que δ e $\epsilon > 0$ são dados. Podemos usar (29) ou (30) para obter um limite inferior para o número necessário de provas para nos garantir que

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \delta\right) \leq \varepsilon.$$

Realmente, de (29) vemos que este será o caso se $p(1-p)/n\delta^2 \leqslant \epsilon$. Resolvendo para n vemos que $n \geqslant p(1-p)/\epsilon\delta^2$. Se usarmos (30), $n \geqslant (4\epsilon\delta^2)^{-1}$ provas serão suficientes. Observamos novamente que estes limites dados pela desigualdade de Chebyshev são pobres e que na realidade um número muito menor de provas pode ser suficiente.

Como uma ilustração da diferença entre estas duas estimativas, escolha $\delta = 0{,}1$ e $\epsilon = 0{,}01$. Então $\delta^2\,\epsilon = 10^{-4}$ e de (30) vemos que para garantir que

$$P\left(\left|\frac{S_n}{n} - p\right| \geq 0{,}1\right) \leq 0{,}01$$

necessitaríamos de $n = 10^4/4 = 2500$ observações. Suponha, entretanto, que soubéssemos que $p = 0{,}1$. Então, como $p(1-p) = 0{,}09$, vemos de (29) que $n \geqslant 0{,}09 \times 10^4 = 900$ observações serão suficientes. Para $p = 1/2$, (29) produz a mesma estimativa de (30), ou seja 2500.

Para ilustrar a pobreza dos limites de Chebyshev no caso da distribuição binomial, suponha que $n = 100$ e $p = 1/2$.

De (29) obtemos então.

$$P\left(\left|\frac{S_n}{n} - 0{,}5\right| \geqslant 0{,}1\right) \leqslant 0{,}25.$$

Este resultado deve ser comparado com o valor exato desta probabilidade que é 0,038.

Exercícios

1. Seja N um número inteiro positivo e seja f a função definida por

$$f(x) = \begin{cases} \dfrac{2x}{N(N+1)}, & x = 1, 2, \ldots, N, \\ 0, & \text{outros valores de } x \end{cases}$$

Mostre que f é uma densidade discreta e obtenha sua média. *Sugestão:*

$$\sum_{x=1}^{N} x = \frac{N(N+1)}{2} \quad \text{e} \quad \sum_{x=1}^{N} x^2 = \frac{N(N+1)(2N+1)}{6}.$$

2. Suponha que X tem uma densidade binomial de parâmetros $n = 4$ e p. Obtenha $E \operatorname{sen}(\pi X/2)$.
3. Suponha que X tem uma densidade de Poisson de parâmetro λ. Determine a média de $(1+X)^{-1}$.
4. Se X tem média 1 e Y tem média 3, qual é a média de $2X + 5Y$?
5. Suponha que X e Y são duas variáveis aleatórias tais que

$$P(|X - Y| \leqslant M) = 1$$

para alguma constante M. Mostre que se Y tem expectância finita, então X tem expectância finita e $|EX - EY| \leqslant M$.

6. Seja X uma variável aleatória com distribuição geométrica e seja $M > 0$ um número positivo. Faça $Z = \min(X, M)$. Determine a média de Z. *Sugestão:* use o Teorema 5.
7. Seja X uma variável aleatória com distribuição geométrica e seja $M > 0$ um número positivo. Faça $Y = \operatorname{Max}(X, M)$. Determine a média de Y. *Sugestão:* determine $P(Y < y)$ e a seguir use o Teorema 5.
8. Suponha que X se distribui uniformemente sobre $\{0, 1, \ldots, N\}$. Determine a média e variância de X usando a sugestão do Exercício 1.
9. Construa um exemplo de uma densidade que tem um momento finito de ordem r, mas não tem nenhum momento de ordem superior a r. *Sugestão:* considere a série $\Sigma_{x=1}^{\infty} x^{-(r+2)}$ e transforme-a em uma densidade.
10. Suponha que X e Y são duas variáveis aleatórias independentes tais que $EX^4 = 2$, $EY^2 = 1$, $EX^2 = 1$, e $EY = 0$. Determine $\operatorname{Var}(X^2 Y)$.
11. Mostre que $\operatorname{Var}(aX) = a^2 \operatorname{Var} X$.
12. Suponha que X se distribui binomialmente com parâmetros n e p. Use a função geratriz de probabilidade de X para determinar sua média e variância.
13. Seja X uma variável aleatória inteira não-negativa.

(a) Mostre que

$$\Phi_X(t) = Et^X, \qquad -1 \leq t \leq 1,$$

$$\Phi'_X(t) = EXt^{X-1}, \qquad -1 < t < 1,$$

e

$$\Phi''_X(t) = EX(X-1)t^{X-2}, \qquad -1 < t < 1.$$

(b) Use o Teorema 4 para demonstrar que se X e Y são variáveis aleatórias inteiras, não-negativas e independentes, então

$$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t), \qquad -1 \leq t \leq 1.$$

14. Sejam X e Y duas variáveis aleatórias independentes tendo segundos momentos finitos. Determine a média e a variância de $2X + 3Y$ em termos das de X e Y.

15. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes tendo uma densidade comum com média μ e variância σ^2. Faça $\bar{X} = (X_1 + \cdots + X_n)/n$.

(a) Escrevendo $X_k - \bar{X} = (X_k - \mu) - (\bar{X} - \mu)$, mostre que

$$\sum_{k=1}^{n} (X_k - \bar{X})^2 = \sum_{k=1}^{n} (X_k - \mu)^2 - n(\bar{X} - \mu)^2.$$

(b) Conclua de (a) que

$$E\left(\sum_{k=1}^{n} (X_k - \bar{X})^2\right) = (n - 1)\sigma^2.$$

16. Suponha que se distribua aleatoriamente n bolas em r caixas. Seja $X_i = 1$, se a caixa i estiver vazia e, seja $X_i = 0$ caso contrário.

(a) Determine EX_i.

(b) Para $i \neq j$ determine $E(X_iX_j)$.

(c) Seja S_r o número de caixas vazias. Escreva $S_r = X_1 + \cdots + X_r$, e use o resultado de (a) para determinar ES_r.

(d) Use os resultados de (a) e (b) para determinar $\operatorname{Var} S_r$.

17. Suponha que tenhamos dois baralhos de n cartas, cada um com as cartas numeradas de 1 a n. Utilizando estas cartas forma-se n pares, cada par contendo uma carta de cada baralho. Dizemos que ocorre um encontro na posição i se o par i é constituído de cartas de mesmo valor. Seja S_n o número de encontros.

(a) Determine ES_n.

(b) Determine $\operatorname{Var} S_n$.

Sugestão: seja $X_i = 1$ se ocorre um encontro na posição i, e seja $X_i = 0$ caso contrário. Então $S_n = X_1 + \cdots + X_n$. Dos resultados do Capítulo 2 sabemos que $P(X_i = 1) = 1/n$ e que se $i \neq j$,

$$P(X_i = 1, X_j = 1) = \frac{1}{n(n - 1)}.$$

18. Considere a variável aleatória S_k introduzida no Exemplo 8. Determine $\operatorname{Var} S_k$.

19. Estabeleça as seguintes propriedades da covariância:

(a) $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$;

(b) $\operatorname{Cov}\left(\sum_{i=1}^{m} a_iX_i, \sum_{j=1}^{n} b_jY_j\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_ib_j \operatorname{Cov}(X_i, Y_j).$

20. Sejam X_1, X_2 e X_3 variáveis aleatórias independentes tendo variâncias finitas e positivas σ_1^2, σ_2^2 e σ_3^2, respectivamente. Obtenha a correlação entre $X_1 - X_2$ e $X_2 + X_3$.

21. Suponha que X e Y são duas variáveis aleatórias tais que $\rho(X, Y) = 1/2$, $\operatorname{Var} X = 1$ e $\operatorname{Var} Y = 2$. Obtenha $\operatorname{Var}(X - 2Y)$.

22. Uma caixa contém 3 bolas vermelhas e 2 pretas. Extrai-se uma amostra sem reposição de tamanho dois. Sejam U e V os números de bolas vermelhas e pretas, respectivamente, na amostra. Determine $\rho(U, V)$.

23. Suponha que uma caixa contém 3 bolas numeradas de 1 a 3. Seleciona-se sem reposição duas bolas da caixa. Seja X o número da primeira bola e Y o número da segunda bola.
Determine $\text{Cov}(X, Y)$ e $\rho(X, Y)$.

24. Suponha que se repete n vezes um experimento tendo r resultados possíveis $1, 2, \ldots, r$ que ocorrem com probabilidades $p_1, \ldots, p_r$. Seja X o número de vezes que ocorre o resultado 1 e seja Y o número de vezes que ocorre o resultado 2.
Mostre que $$\rho(X, Y) = -\sqrt{\frac{p_1 p_2}{(1 - p_1)(1 - p_2)}},$$
através das seguintes etapas. Faça $I_i = 1$ se ocorre resultado 1 na i-ésima repetição, e $I_i = 0$, caso contrário. Similarmente faça $J_i = 1$ se ocorre resultado 2 na i-ésima repetição, $J_i = 0$ para outros valores. Então $X = I_1 + \cdots + I_n$ e $Y = J_1 + \cdots + J_n$. A seguir mostre que

(a) $E(I_i J_i) = 0$.

(b) Se $i \neq j, E(I_i J_j) = p_1 p_2$.

(c) $$E(XY) = E\left(\sum_{i=1}^{n} I_i J_i\right) + E\left(\sum_{i=1}^{n} \sum_{j \neq i} I_i J_j\right) = n(n - 1) p_1 p_2.$$

(d) $\text{Cov}(X, Y) = -n p_1 p_2$.

(e) $$\rho(X, Y) = -\sqrt{\frac{p_1 p_2}{(1 - p_1)(1 - p_2)}}.$$

25. Suponha que uma população de r objetos é constituída de r_1 objetos de tipo 1, r_2 objetos do tipo 2 e r_3 objetos do tipo 3, onde $r_1 + r_2 + r_3 = r$. Toma-se desta população uma amostra aleatória sem reposição de tamanho $n \leqslant r$. Seja X o número de objetos do tipo 1 e Y o número de objetos do tipo 2 na amostra. Determine $\rho(X, Y)$ da seguinte forma: Faça $I_i = 1$ ou 0 dependendo de se o i-ésimo objeto é do tipo 1 ou não, e $J_i = 1$ ou 0 dependendo de se o i-ésimo objeto é do tipo 2 ou não.

(a) Mostre que $EI_i = r_1/r$ e $EJ_i = r_2/r$.

(b) Mostre que para $i \neq j$,
$$EI_i J_j = \frac{r_1 r_2}{r(r - 1)}$$
e que $E(I_i J_i) = 0$.

(c) Faça $X = I_1 + \cdots + I_n$, e $Y = J_1 + \cdots + J_n$ e use a e b, para determinar $E(XY)$, $\text{Var}\, X$ e $\text{Var}\, Y$.

(d) Use (c) para determinar $\rho(X, Y)$. Compare com o correspondente coeficiente de correlação do Exercício 24 $\rho_1 = r_1/r$ e $p_2 = r_2/r$.

26. Seja X uma variável aleatória tendo a densidade f dada por

$$f(x) = \begin{cases} 1/18, & x = 1, 3, \\ 16/18, & x = 2. \end{cases}$$

Mostre que existe um valor de δ tal que $P(|X - \mu| \geqslant \delta) = \text{Var } X/\delta^2$, de modo que em geral o limite dado pela desigualdade de Chebyshev não pode ser melhorado.

27. Um fabricante de parafusos sabe que 5% de sua produção é defeituosa. Ele oferece uma garantia sobre sua remessa de 10.000 itens, prometendo reembolsar o dinheiro se mais de a parafusos forem defeituosos. Qual o menor valor que o fabricante pode atribuir a a e ainda continuar seguro de que não precisa reembolsar o dinheiro em mais de 1% de vezes?

28. Suponha que X tem densidade de Poisson de parâmetro λ. Use a desigualdade de Chebyshev para verificar as seguintes desigualdades:

(a) $P\left(X \leq \dfrac{\lambda}{2}\right) \leq \dfrac{4}{\lambda}$; (b) $P(X \geq 2\lambda) \leq \dfrac{1}{\lambda}$.

29. Seja X uma variável aleatória inteira não-negativa cuja função geratriz de probabilidade $\Phi_X(t) = Et^X$ é finita para todo t e seja x_0 um número inteiro positivo. Argumentando como na demonstração da desigualdade de Chebyshev, verifique as seguintes desigualdades:

(a) $P(X \leq x_0) \leq \dfrac{\Phi_X(t)}{t^{x_0}}, \quad 0 \leq t \leq 1;$

(b) $P(X \geq x_0) \leq \dfrac{\Phi_X(t)}{t^{x_0}}, \quad t \geq 1.$

30. Suponha que X tem densidade de Poisson de parâmetro λ. Verifique as seguintes desigualdades:

(a) $P\left(X \leq \dfrac{\lambda}{2}\right) \leq \left(\dfrac{2}{e}\right)^{\lambda/2}$; (b) $P(X \geq 2\lambda) \leq \left(\dfrac{e}{4}\right)^{\lambda}$.

Sugestão: use cálculo para minimizar os segundos membros das desigualdades do Exercício 29. Estas desigualdades são muito mais rigorosas do que as dadas no Exercício 28.

Os Exercícios 31 e 36 desenvolvem e aplicam as noções de densidade condicional. Sejam X e Y duas variáveis aleatórias discretas. Define-se a *densidade condicional* $f_{Y|X}(y \mid x)$ *de* Y *dado* $X = x$ *através de*

$$f_{Y|X}(y \mid x) = \begin{cases} P(Y = y \mid X = x), & \text{se } P(X = x) > 0, \\ 0, & \text{para outros valores de } x. \end{cases}$$

Segue-se que $f_{Y|X}(y \mid x)$ é uma densidade em y para qualquer x tal que $P(X = x) > 0$. O Exemplo 14(d) do Capítulo 3 pode ser interpretado dizendo que se X e Y são independentes e se distribuem geometricamente com parâmetro p, então, para $z \geqslant 0$, a densidade condicional de Y dado $X + Y = z$ é a densidade uniforme em $\{0, 1, \ldots, z\}$.

Suponha que Y tem expectância finita. Define-se como *expectância condicional de Y dado $x = x$* a média da densidade condicional de Y dado $X = x$, isto é,

$$E[Y \mid X = x] = \sum_y y f_{Y|X}(y \mid x).$$

31. Verifique as seguintes propriedades de densidade condicional e da expectância condicional:

(a) $f_Y(y) = \sum_x f_X(x) f_{Y|X}(y \mid x)$; (b) $EY = \sum_x f_X(x) E[Y \mid X = x]$.

32. Sejam X e Y duas variáveis aleatórias independentes com densidade geométrica de parâmetro p. Determine $E[Y \mid X + Y = z]$ onde z é um inteiro não-negativo. *Sugestão:* Use o Exemplo 14(d) e o Exercício 8.

33. Sejam X e Y duas variáveis aleatórias independentes com distribuições de Poisson de parâmetros λ_1 e λ_2, respectivamente. Determine $E[Y \mid X + Y = z]$ onde z é um inteiro não-negativo. *Sugestão:* use o resultado do Exercício 35 do Capítulo 3.

34. Seja N uma variável aleatória inteira não-negativa. Sejam $\{Y_n\}$, $n \geqslant 0$, variáveis aleatórias com expectâncias finitas e independentes de N. Mostre que

$$E[Y_N \mid N = n] = EY_n.$$

35. Sejam $\{X_n\}$, $n \geqslant 1$, variáveis aleatórias independentes tendo média μ e variância σ^2 finitas e comuns. Faça $S_0 = 0$ e $S_n = X_1 + \cdots + X_n, n \geqslant 1$. Seja N uma variável aleatória inteira não-negativa com média e variância finitas, e suponha que N é independente de todas variáveis aleatórias definidas em termos de $\{X_n\}$, $n \geqslant 1$. Então S_n tem média e variância finitas. Mostre que

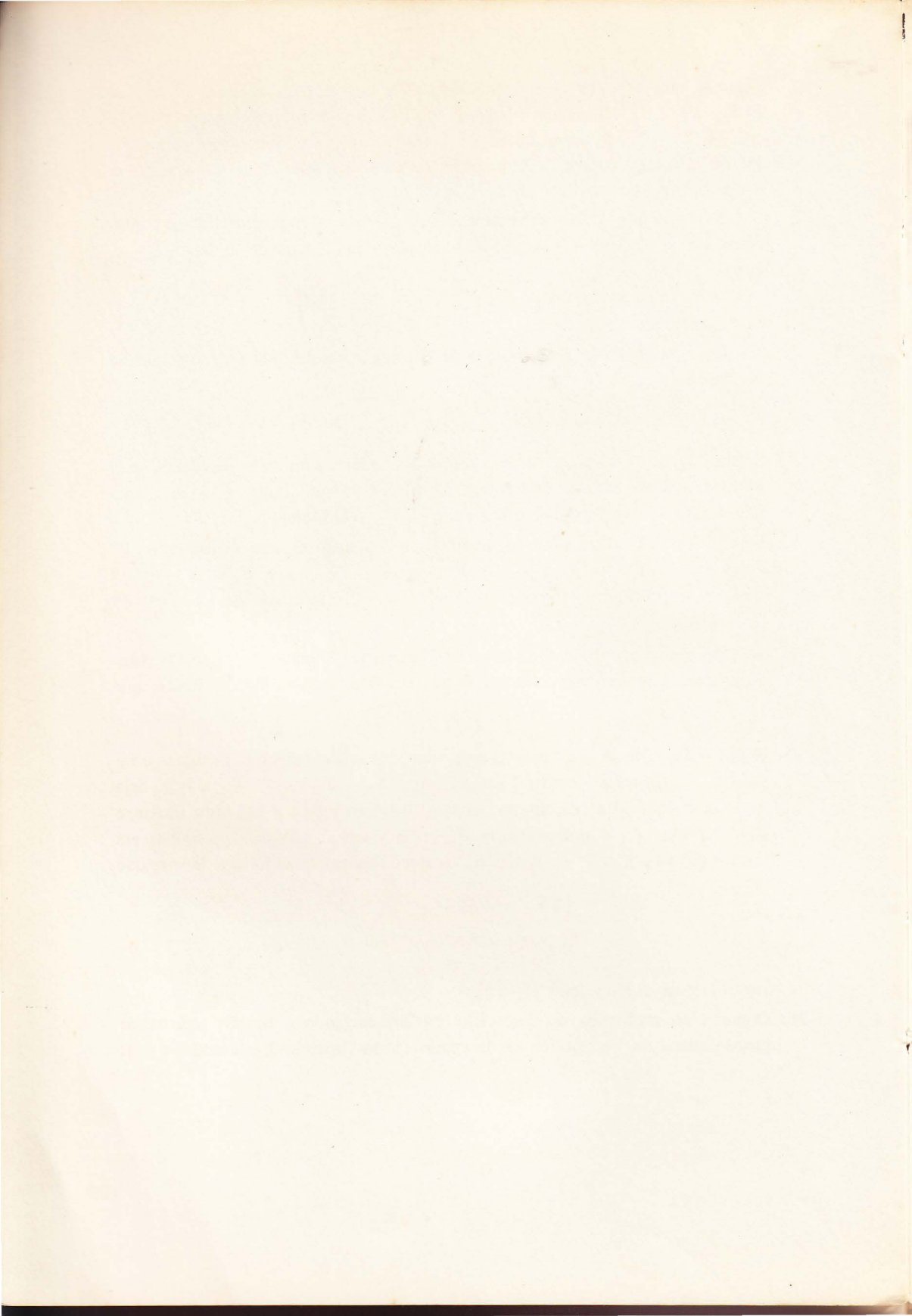
$$ES_N = \mu EN, \qquad ES_N^2 = \sigma^2 EN + \mu^2 EN^2,$$

e

$$\operatorname{Var} S_N = \sigma^2 EN + \mu^2 \operatorname{Var} N.$$

Sugestão: use os Exercícios 31(b) e 34.

36. Obtenha os resultados do Exercício 35 diferenciando a função geratriz de probabilidade de S_N obtida no Exemplo 19 do Capítulo 3 e fazendo $t = 1$.

5

VARIÁVEIS ALEATÓRIAS CONTÍNUAS

No Capítulo 3 consideramos as variáveis aleatórias discretas e suas densidades como, por exemplo, a binomial, a hipergeométrica e a de Poisson. Nas aplicações, estas variáveis aleatórias representam tipicamente o número de objetos de um certo tipo, como o número de bolas vermelhas em uma amostra aleatória de tamanho n, com ou sem reposição, ou o número de chamadas que chegam a uma central telefônica num minuto.

Existem muitas situações, tanto teóricas como aplicadas, em que as variáveis aleatórias naturais a considerar são "contínuas" em vez de discretas. Como aproximação inicial, podemos definir uma variável aleatória contínua X num espaço de probabilidade Ω como uma função $X(\omega)$, $\omega \in \Omega$, tal que

$$P(\{\omega \mid X(\omega) = x\}) = 0, \qquad -\infty < x < \infty,$$

isto é, tal que X assume qualquer valor específico x com probabilidade zero.

É fácil pensar em exemplos de variáveis aleatórias contínuas. Como primeira ilustração, considere um modelo probabilístico para os tempos de desintegração de um número finito de partículas radioativas. Seja T o tempo que decorre até a desintegração da primeira partícula. Então, T seria uma variável aleatória contínua, pois a probabilidade de que a primeira desintegração ocorra num tempo específico (por exemplo, $T = 2.0000\ldots$ segundos) é zero. Como segundo exemplo, considere o experimento de escolher ao acaso um ponto de um subconjunto S do espaço euclidiano n-dimensional, tendo um volume n-dimensional finito não-nulo (lembre-se da discussão deste experimento no Capítulo 1). Seja X a variável aleatória que representa a primeira coordenada do ponto escolhido. Suponha, por exemplo, que $n = 2$ e que S seja um disco de raio unitário no plano, centrado na origem. Então, o conjunto de pontos em S que tem a primeira coordenada zero é um segmento de reta no plano. Qualquer segmento como este tem *área* zero e portanto probabilidade zero.

De uma forma geral, as variáveis aleatórias que representam medidas de grandezas físicas como coordenadas espaciais, peso, tempo, temperatura e voltagem

são descritas mais adequadamente como variáveis aleatórias contínuas. Variáveis aleatórias associadas às contagens de objetos ou eventos são exemplos típicos de variáveis aleatórias discretas.

Entretanto, existem casos em que tanto a formulação discreta como a contínua poderiam ser apropriadas. Assim, embora normalmente a medida de comprimento seja considerada como uma variável aleatória contínua, poderíamos considerar a medida arredondada para um certo número de casas decimais, e portanto como sendo uma variável aleatória discreta.

5.1. VARIÁVEIS ALEATÓRIAS E SUAS FUNÇÕES DE DISTRIBUIÇÃO

Nas aplicações, uma variável aleatória representa uma quantidade numérica definida em termos do resultado de um experimento aleatório. Matematicamente, entretanto, uma variável aleatória X é uma função real definida num espaço de probabilidade. Naturalmente desejamos que $P(X \leqslant x)$ seja bem definida para todo número real x. Em outras palavras, se $(\Omega, \mathscr{A}, P)$ for um espaço de probabilidade sobre o qual se define X, desejamos que

$$\{\omega \mid X(\omega) \leqslant x\}$$

seja um evento (isto é, um elemento de $\mathscr{A}$). Isto nos leva às definições abaixo.

Definição 1. Uma variável aleatória X num espaço de probabilidade $(\Omega, \mathscr{A}, P)$ é uma função real $X(\omega)$, $\omega \in \Omega$, tal que $\{\omega \mid X(\omega) \leqslant x\}$ é um evento para $-\infty < x < \infty$.

Definição 2. A função de distribuição F de uma variável aleatória X é a função

$$F(x) = P(X \leqslant x), \qquad -\infty < x < \infty.$$

A função de distribuição é útil na determinação de diferentes probabilidades associadas com a variável aleatória X. Um exemplo disto é a fórmula

$$P(a < X \leqslant b) = F(b) - F(a), \qquad a \leqslant b. \tag{1}$$

Para verificar a validade de (1), faça-se $A = \{\omega \mid X(\omega) \leqslant a\}$ e $B = \{\omega \mid X(\omega) \leqslant b\}$. Então $A \subseteq B$ e, segundo a definição de variável aleatória, A e B são ambos eventos. Portanto $\{\omega \mid a < X \leqslant b\} = B \cap A^c$ é um evento e (1) é um caso especial do fato demonstrado na Seção 1.3, de acordo com o qual, se $A \subseteq B$, então

$$P(B \cap A^c) = P(B) - P(A).$$

Exemplo 1. Considere o experimento de escolher ao acaso um ponto do disco de raio R no plano, centrado na origem. Para tornar o experimento mais interessante, podemos imaginá-lo como sendo o resultado do lançamento de um dardo sobre um alvo circular. Associado a este experimento, está o espaço uniforme de

probabilidade descrito na Seção 1.2. Seja X a variável aleatória que representa a distância do ponto escolhido à origem. É fácil determinar a função de distribuição de X. Se $0 \leqslant x \leqslant R$, o evento $\{\omega \mid X(\omega) \leqslant x\}$ é o disco de raio x no plano, centrado na origem. Sua área é πx^2. Assim, pela definição de um espaço uniforme de probabilidade,

$$P(X \leqslant x) = \frac{\pi x^2}{\pi R^2} = \frac{x^2}{R^2}, \quad 0 \leqslant x \leqslant R.$$

Se, $x < 0$, então $P(X \leqslant x) = 0$. Se $x > R$, então $P(X \leqslant x) = 1$. Assim, a função de distribuição F da variável aleatória X é dada por

$$(2) \qquad F(x) = \begin{cases} 0, & x < 0, \\ x^2/R^2, & 0 \leqslant x \leqslant R, \\ 1, & x > R. \end{cases}$$

O gráfico de F é dado na Figura 1. Segue-se das fórmulas (1) e (2) que se $0 \leqslant a \leqslant b \leqslant R$, então

$$P(a < X \leqslant b) = F(b) - F(a) = \frac{b^2 - a^2}{R^2}.$$

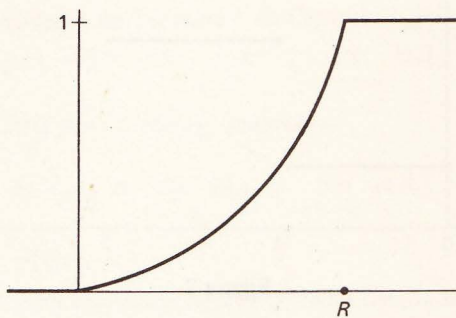



Figura 1

Exemplo 2. Considere um modelo probabilístico para os tempos de desintegração de um número finito de partículas radioativas. Seja X o tempo de desintegração de uma partícula específica. Obtenha a função de distribuição de X.

Como vimos na Seção 1.1, para um valor positivo adequado de λ,

$$P(a < X \leqslant b) = e^{-\lambda a} - e^{-\lambda b}, \quad 0 \leqslant a \leqslant b < \infty.$$

Já que X assume somente valores positivos, $P(X \leqslant x) = 0$ para $x \leqslant 0$ e, em particular, $P(X \leqslant 0) = 0$. Para $0 < x < \infty$,

$$P(X \leqslant x) = P(X \leqslant 0) + P(0 < X \leqslant x) =$$
$$= P(0 < X \leqslant x) =$$
$$= 1 - e^{-\lambda x}.$$

Assim, X tem a função de distribuição F dada por

(3) $$F(x) = \begin{cases} 0, & x \leqslant 0, \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

Naturalmente as variáveis aleatórias discretas também tem funções de distribuição, duas das quais foram determinadas nos Exemplos 10 e 11 do Capítulo 3.

Exemplo 3. Suponha que X tenha uma distribuição binomial de parâmetros $n = 2$ e $p = 1/2$. Então $f(0) = 1/4$, $f(1) = 1/2$ e $f(2) = 1/4$. Conseqüentemente

$$F(x) = \begin{cases} 0, & x < 0, \\ 1/4, & 0 \leqslant x < 1, \\ 3/4, & 1 \leqslant x < 2, \\ 1, & 2 \leqslant x. \end{cases}$$

O gráfico desta função de distribuição é dado na Figura 2.

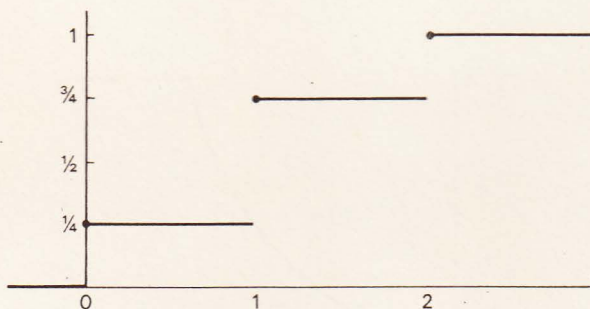



Figura 2

5.1.1. PROPRIEDADES DE FUNÇÕES DE DISTRIBUIÇÃO. Nem todas as funções ocorrem como funções de distribuição, pois as últimas devem satisfazer certas condições. Seja X uma variável aleatória e seja F sua função de distribuição. Então

(i) $0 \leqslant F(x) \leqslant 1$ *para todo* x.

(ii) F *é uma função não decrescente de* x.

A propriedade (i) decorre imediatamente da propriedade de definição $F(x) = P(X \leqslant x)$. Para verificar a validade de (ii), precisamos simplesmente observar que se $x < y$, então

$$F(y) - F(x) = P(x < X \leqslant y) \geqslant 0.$$

Diz-se que uma função f tem um limite L à direita (à esquerda) no ponto x se $f(x + h) \to L$ para $h \to 0$ através de valores positivos (negativos). Quando existem, representam-se os limites à direita e à esquerda por $f(x +)$ e $f(x -)$. Não é difícil mostrar que se f for limitada e não-crescente ou não-decrescente, então $f(x +)$ e $f(x -)$ existem para todo x. Sob as mesmas condições, f tem os limites $f(-\infty)$ para $x \to -\infty$ e $f(+\infty)$ para $x \to +\infty$.

Das propriedades (i) e (ii) e da discussão do parágrafo precedente, segue-se que a função de distribuição F tem os limites $F(x +)$ e $F(x -)$ para todo x, bem como os limites $F(-\infty)$ e $F(+\infty)$.

(iii) $F(-\infty) = 0$ *e* $F(+\infty) = 1$.

(iv) $F(x+) = F(x)$ *para todo* x.

Para avaliar $F(-\infty)$ e $F(+\infty)$ precisamos apenas obter os limites de $F(n)$ para $n \to -\infty$ e $n \to +\infty$. (Isto decorre do fato de que F é não-decrescente). Seja

$$B_n = \{ \omega \mid X(\omega) \leqslant n \}.$$

Então $\cdots \subseteq B_{-2} \subseteq B_{-1} \subseteq B_0 \subseteq B_1 \subseteq B_2 \subseteq \cdots$. Também

$$\bigcap_{n=0}^{-\infty} B_n = \varnothing \quad \text{e} \quad \bigcup_{n=0}^{+\infty} B_n = \Omega.$$

Segue-se então os resultados do Teorema 1 do Capítulo 1 que

$$\lim_{n \to -\infty} P(B_n) = P(\varnothing) = 0 \quad \text{e} \quad \lim_{n \to +\infty} P(B_n) = P(\Omega) = 1.$$

De vez que $F(n) = P(X \leqslant n) = P(B_n)$, vemos que

$$F(-\infty) = \lim_{n \to -\infty} F(n) = \lim_{n \to -\infty} P(B_n) = 0$$

e analogamente que $F(+\infty) = 1$.

A propriedade (iv) estabelece que F é uma função contínua à direita e que

$$F(x+) = P(X \leqslant x), \quad -\infty < x < \infty. \tag{4}$$

Um resultado intimamente relacionado com (4) é

$$F(x-) = P(X < x), \quad -\infty < x < \infty. \tag{5}$$

As demonstrações de (4) e (5) são semelhantes à de (iii). Para demonstrar (4), por exemplo, precisamos apenas mostrar que $F(x + 1/n) \to P(X \leqslant x)$ para $n \to +\infty$. Isto pode ser feito definindo

$$B_n = \left\{ \omega \mid X(\omega) \leqslant x + \frac{1}{n} \right\},$$

observando que $\bigcap_n B_n = \{\omega \mid X(\omega) \leq x\}$ e repetindo o argumento de (iii).

Vemos imediatamente de (4) e (5) que

$$(6) \qquad F(x+) - F(x-) = P(X = x), \qquad -\infty < x < \infty.$$

Esta fórmula estabelece que se $P(X = x) > 0$, então F apresenta um salto de magnitude $P(X = x)$ no ponto x. Se $P(X = x) = 0$, então F é contínua no ponto x. Lembramos o conceito de uma variável aleatória contínua (introdução deste capítulo).

Definição 3. Chama-se uma variável aleatória X de variável aleatória contínua se

$$P(X = x) = 0, \qquad -\infty < x < \infty.$$

Vemos agora que X é uma variável aleatória contínua se, e somente se, sua função de distribuição F for contínua para cada x, isto é, F for uma função contínua. Se X for uma variável aleatória contínua, então além de (1) temos que

$$(7) \qquad \begin{aligned} P(a < X < b) &= P(a \leqslant X \leqslant b) = P(a \leqslant X < b) \\ &= F(b) - F(a), \end{aligned}$$

de modo que $<$ e $\leqslant$ podem ser usados indiscriminadamente neste contexto. As diversas propriedades de uma função de distribuição estão ilustradas na Figura 3. (Observe que a variável aleatória com esta função de distribuição não seria nem discreta nem contínua.)

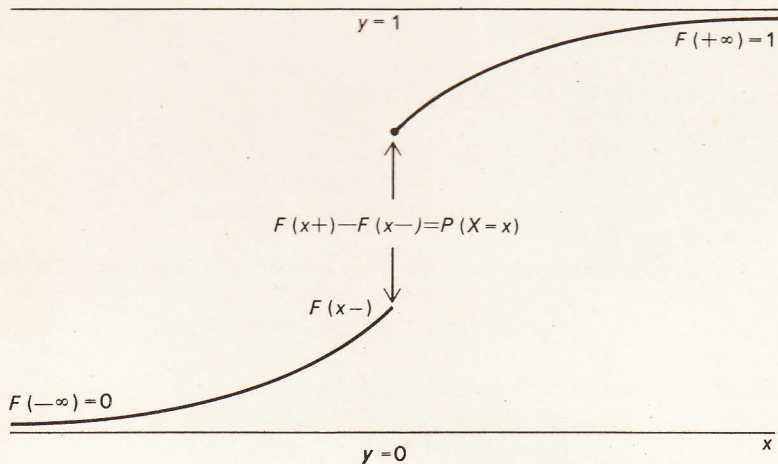



Figura 3

Considere a variável aleatória X definida no Exemplo 1. Da fórmula (2) ou Figura 1 vemos que sua distribuição é contínua. Então X é uma variável aleatória contínua. Da mesma forma torna-se claro de (3) que a variável aleatória do Exemplo 2 é uma variável aleatória contínua.

A maioria das variáveis aleatórias que ocorrem nas aplicações práticas são discretas ou contínuas. Existem algumas exceções. Considere o Exemplo 2. Neste exemplo, X representa o tempo até a desintegração de uma partícula específica. Se o experimento durar apenas um tempo especificado, digamos até o tempo $t_0 > 0$, e a partícula não se desintegrar até este instante, o seu tempo real até a desintegração não será observado. Uma possível forma de contornar esta dificuldade é definir uma nova variável aleatória Y da seguinte forma

$$Y(\omega) = \begin{cases} X(\omega) & \text{se} \quad X(\omega) \leqslant t_0, \\ t_0 & \text{se} \quad X(\omega) > t_0. \end{cases}$$

Assim Y é o tempo de degradação, se este tempo for observado (isto é, for menor ou igual a t_0) e $Y = t_0$ caso contrário. A função de distribuição F_Y de Y é dada por

$$F_Y(y) = \begin{cases} 0, & y < 0, \\ 1 - e^{-\lambda y}, & 0 \leqslant y < t_0, \\ 1, & y \geqslant t_0. \end{cases}$$

A função de distribuição apresenta um salto de magnitude $e^{-\lambda t_0}$ no ponto $y = t_0$. Assim fica claro que a variável aleatória Y que acabamos de construir não é discreta nem contínua.

Nós definimos funções de distribuição em termos de variáveis aleatórias. Elas podem ser definidas diretamente.

Definição 4. Uma função de distribuição é qualquer função F que satisfaz as propriedades (i)-(iv); isto é,

(i) $0 \leqslant F(x) \leqslant 1$ para todo x,
(ii) F é uma função não decrescente de x,
(iii) $F(-\infty) = 0$ e $F(+\infty) = 1$,
(iv) $F(x+) = F(x)$ para todo x.

Em textos mais avançados demonstra-se que se F *for uma função de distribuição, necessariamente existe um espaço de probabilidade e uma variável aleatória* X *definida neste espaço tal que* F *é a função de distribuição de* X.

5.2. DENSIDADES DE VARIÁVEIS ALEATÓRIAS CONTÍNUAS

Na prática define-se geralmente as funções de distribuição em termos de funções de densidade.

Definição 5. Uma função de densidade (em relação à integração) é uma função não-negativa f tal que

$$\int_{-\infty}^{\infty} f(x)\,dx = 1.$$

Observe que se f for uma função de densidade, então a função F definida por

(8) $$F(x) = \int_{-\infty}^{x} f(y)\, dy, \qquad -\infty < x < \infty,$$

é uma função contínua que satisfaz as propriedades (i)-(iv) da Seção 5.1.1. Assim (8) define uma função de distribuição contínua. Dizemos que esta função de distribuição tem densidade f. É possível, embora difícil, construir exemplos de funções de distribuição contínuas que não possuem densidades. Aquelas que realmente possuem densidades são chamadas de funções de distribuição *absolutamente contínuas.*

Se X for uma variável aleatória contínua tendo F como sua função de distribuição, onde F é dada por (8), então f é também chamada de densidade de X. Usaremos o termo "função de densidade" em relação a funções de densidade discreta ou a funções de densidade em relação à integração. Deverá ficar claro, no contexto, o tipo de função de densidade que está sendo considerado. Por exemplo, a frase "seja X uma variável aleatória contínua de densidade f" implica necessariamente que f é uma função de densidade em relação à integração.

Segue-se de (1) e (8) que se X for uma variável aleatória contínua de densidade f, então

(9) $$P(a \leq X \leq b) = \int_{a}^{b} f(x)\, dx, \qquad a \leq b,$$

ou de uma forma um pouco mais geral, que

(10) $$P(X \in A) = \int_{A} f(x)\, dx$$

se A for a união de um número finito ou infinito contável de intervalos disjuntos. Assim $P(X \in A)$ pode ser representada como a área sob a curva f com x abrangendo o conjunto A (ver Figura 4).

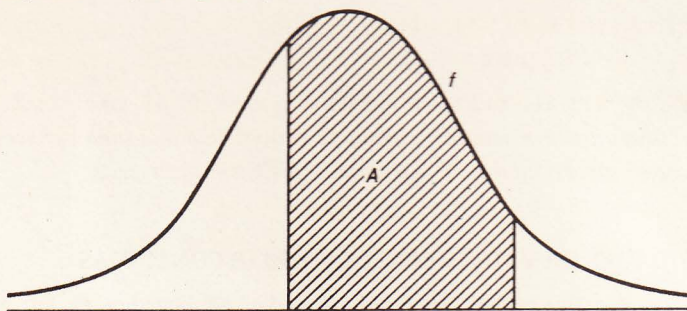



Figura 4

Na maioria das aplicações, a maneira mais fácil de determinar as densidades de variáveis aleatórias contínuas é diferenciar (8) e obter

(11) $$f(x) = F'(x), \qquad -\infty < x < \infty.$$

Estritamente falando, (11) é válida em todos os pontos x onde f é contínua.

Exemplo 4 Seja X a variável aleatória do Exemplo 1 tendo a função de densidade F dada por (2). Então

$$(12) \qquad F'(x) = \begin{cases} 0, & x < 0, \\ 2x/R^2, & 0 \leqslant x < R, \\ 0, & x > R. \end{cases}$$

A função F não é diferenciável no ponto $x = R$. Entretanto, se definirmos f por $f(x) = F'(x), x \neq R$, e $f(R) = 0$, este f será uma densidade para F.

Observamos que (8) não define f de maneira única, uma vez que podemos alterar o valor de uma função em um número finito de pontos sem alterar a integral da função sobre intervalos. Uma forma típica de definir f é fazer $f(x) = F'(x)$ sempre que $F'(x)$ existe e $f(x) = 0$ caso contrário. Isto define uma densidade de F desde que F seja sempre contínua e que F' exista e seja contínua em todos, exceto num número finito de pontos.

Existem outras maneiras de derivar ou verificar fórmulas para a densidade de uma função de distribuição contínua F. Dada uma função de densidade f, podemos mostrar que f é uma função de densidade de F verificando que (8) é válida. Alternativamente, podemos inverter este processo e mostrar que F pode ser escrita na forma (8) para alguma função não-negativa f. Então f é necessariamente uma densidade de F. Estes métodos, essencialmente equivalentes entre si, são geralmente mais complicados que o da diferenciação. Entretanto são mais rigorosos e evitam a necessidade de consideração especial dos pontos em que $F'(x)$ não existe.

Ilustraremos o uso destes métodos no primeiro exemplo da subseção seguinte.

5.2.1. FÓRMULAS PARA MUDANÇAS DE VARIÁVEL. Seja X uma variável aleatória contínua de densidade f. Discutimos métodos de obter a densidade de uma variável aleatória Y que seja função de X.

Exemplo 5. Seja X uma variável aleatória contínua de densidade f. Obtenha a densidade da variável aleatória $Y = X^2$.

Para resolver este problema, inicialmente representamos por F e G as funções de distribuição de X e Y, respectivamente. Então $G(y) = 0$ para $y \leqslant 0$. Para $y > 0$

$$\begin{aligned} G(y) &= P(Y \leqslant y) = P(X^2 \leqslant y) \\ &= P(-\sqrt{y} \leqslant X \leqslant \sqrt{y}) \\ &= F(\sqrt{y}) - F(-\sqrt{y}) \end{aligned}$$

e por diferenciação vemos que

$$G'(y) = \frac{1}{2\sqrt{y}} (F'(\sqrt{y}) + F'(-\sqrt{y}))$$

$$= \frac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(-\sqrt{y})).$$

Assim $Y = X^2$ tem a densidade g dada por

$$(13) \qquad g(y) = \begin{cases} \dfrac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(-\sqrt{y})) & \text{para} \quad y > 0, \\ 0 & \text{para} \quad y \leqslant 0. \end{cases}$$

Embora (13) seja válido em geral, nossa derivação depende de diferenciação que pode não ser válida para todos os pontos. Para dar uma demonstração elementar porém rigorosa de (13), podemos definir g através do segundo membro de (13) e escrever para $x > 0$

$$\int_{-\infty}^{x} g(y)\, dy = \int_{0}^{x} \frac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(-\sqrt{y}))\, dy.$$

Fazendo a mudança de variável $z = \sqrt{y}$ (de modo que $dz = dy/2\sqrt{y}$), obtemos

$$\int_{-\infty}^{x} g(y)\, dy = \int_{0}^{\sqrt{x}} (f(z) + f(-z))\, dz$$

$$= \int_{-\sqrt{x}}^{\sqrt{x}} f(z)\, dz$$

$$= F(\sqrt{x}) - F(-\sqrt{x}) = G(x),$$

de modo que g é geralmente uma densidade de G.

Daqui para frente usaremos livremente a diferenciação para estabelecer fórmulas como (13), sabendo que se necessário poderíamos apresentar derivações alternativas via integração.

Consideremos agora o uso de (13) para obter a densidade de X^2, onde X é a variável aleatória definida no Exemplo 1. Determinamos no Exemplo 4 a densidade de X como sendo $f(x) = 2x/R^2$ para $0 \leqslant x < R$ e $f(x) = 0$ para outros valores de x. Assim, usando (13) vemos que X^2 tem a densidade g dada por

$$g(y) = \frac{1}{2\sqrt{y}} \frac{2\sqrt{y}}{R^2} = \frac{1}{R^2}, \quad 0 < y < R^2,$$

e $g(y) = 0$ para outros valores de y. Esta densidade é uma densidade uniforme em $(0, R^2)$ de acordo com a seguinte.

Definição 6. Sejam a e b duas constantes com $a < b$. A densidade uniforme no intervalo (a, b) é a densidade f definida por

$$(14) \qquad f(x) = \begin{cases} (b-a)^{-1} & \text{para } a < x < b, \\ 0 & \text{para outros valores de } x. \end{cases}$$

A função de distribuição correspondente a (14) é dada por

$$(15) \qquad F(x) = \begin{cases} 0, & x < a, \\ (x-a)/(b-a), & a \leqslant x \leqslant b, \\ 1, & x > b. \end{cases}$$

Não é difícil encontrar outros exemplos de variáveis aleatórias uniformemente distribuídas. Quando se gira um dial bem equilibrado, é razoável supor que o ângulo do ponteiro (convenientemente definido em radianos) ao parar, após um grande número de revoluções, se distribua uniformemente em $(-\pi, \pi)$ ou equivalentemente em $(0,2\pi)$. Nas aplicações da teoria da probabilidade em cálculo numérico, supõe-se freqüentemente que o erro de arredondamento cometido pelo abandono de todos os dígitos além de n casas decimais se distribui uniformemente em $(0, 10^{-n})$.

Exemplo 6. Suponha que X se distribui uniformemente em (0,1). Obtenha a densidade de $Y = -\lambda^{-1} \log(1 - X)$ para $\lambda > 0$.

Seja G a função de distribuição de Y. Observamos inicialmente que Y é uma variável aleatória positiva e conseqüentemente $G(y) = 0$ para $y \leqslant 0$. Para $y > 0$ temos

$$\begin{aligned} G(y) = P(Y \leqslant y) &= P(-\lambda^{-1} \log(1 - X) \leqslant y) \\ &= P(\log(1 - X) \geqslant -\lambda y) \\ &= P(1 - X \geqslant e^{-\lambda y}) \\ &= P(X \leqslant 1 - e^{-\lambda y}) \\ &= 1 - e^{-\lambda y}. \end{aligned}$$

Portanto $G'(y) = \lambda e^{-\lambda y}$ para $y > 0$ e $G'(y) = 0$ para $y < 0$. Portanto a densidade de Y é dada por

$$(16) \qquad g(y) = \begin{cases} \lambda e^{-\lambda y}, & y > 0, \\ 0, & y \leqslant 0. \end{cases}$$

Esta densidade chama-se *densidade exponencial de parâmetro* λ e será discutida mais detalhadamente na seção seguinte.

O exemplo acima é um caso especial de problemas que podem ser resolvidos por meio do teorema seguinte.

Teorema 1. Seja φ uma função diferenciável, estritamente crescente ou estritamente decrescente em um intervalo I, e sejam $\varphi(I)$ o contradomínio de φ e φ^{-1} a função inversa de φ. Seja X uma variável aleatória contínua de densidade f tal que $f(x) = 0$ para $x \notin I$. Então $Y = \varphi(X)$ tem densidade g dada por $g(y) = 0$ para $y \notin \varphi(I)$ e

$$(17) \qquad g(y) = f(\varphi^{-1}(y)) \left| \frac{d}{dy} \varphi^{-1}(y) \right|, \quad y \in \varphi(I).$$

É um pouco mais sugestivo escrever (17) em sua forma equivalente

$$(18) \qquad g(y) = f(x) \left| \frac{dx}{dy} \right|, \quad y \in \varphi(I) \quad \text{e} \quad x = \varphi^{-1}(y)$$

(ou alternativamente $g(y)|dy| = f(x)|dx|$).

Para derivar (17), sejam F e G as funções de distribuição de X e Y respectivamente. Suponha inicialmente que φ seja estritamente crescente (isto é, $\varphi(x_1) < \varphi(x_2)$ se $x_1 < x_2$, $x_1 \in I$ e $x_2 \in I$). Então φ^{-1} é estritamente crescente em $\varphi(I)$ e para $y \in \varphi(I)$,

$$\begin{aligned} G(y) &= P(Y \leqslant y) \\ &= P(\varphi(X) \leqslant y) \\ &= P(X \leqslant \varphi^{-1}(y)) \\ &= F(\varphi^{-1}(y)). \end{aligned}$$

Assim, pela regra de diferenciação em cadeia

$$\begin{aligned} G'(y) &= \frac{d}{dy} F(\varphi^{-1}(y)) \\ &= F'(\varphi^{-1}(y)) \frac{d}{dy} \varphi^{-1}(y) \\ &= f(\varphi^{-1}(y)) \frac{d}{dy} \varphi^{-1}(y). \end{aligned}$$

Mas

$$\frac{d}{dy} \varphi^{-1}(y) = \left| \frac{d}{dy} \varphi^{-1}(y) \right|$$

porque φ^{-1} é estritamente crescente, de modo que (17) é válida. Suponha a seguir que φ seja estritamente decrescente em I. Então φ^{-1} é estritamente decrescente em $\varphi(I)$, e para $y \in \varphi(I)$.

$$\begin{aligned} G(y) &= P(Y \leqslant y) \\ &= P(\varphi(X) \leqslant y) \\ &= P(X \geqslant \varphi^{-1}(y)) \\ &= 1 - F(\varphi^{-1}(y)). \end{aligned}$$

Assim

$$\begin{aligned} G'(y) &= -F'(\varphi^{-1}(y)) \frac{d}{dy} \varphi^{-1}(y) \\ &= f(\varphi^{-1}(y)) \left(- \frac{d}{dy} \varphi^{-1}(y) \right). \end{aligned}$$

Mas

$$- \frac{d}{dy} \varphi^{-1}(y) = \left| \frac{d}{dy} \varphi^{-1}(y) \right|$$

porque φ^{-1} é estritamente decrescente. Portanto vemos que G tem a densidade g dada por (17) em qualquer um dos casos.

Exemplo 7. Seja X uma variável aleatória tendo uma densidade exponencial de parâmetro λ. Obtenha a densidade de $Y = X^{1/\beta}$, onde $\beta \neq 0$.

De acordo com a definição do exemplo anterior, a densidade de X é dada por $f(x) = \lambda e^{-\lambda x}$ para $x > 0$ e $f(x) = 0$ para $x \leqslant 0$. O teorema acima é aplicável com $\varphi(x) = x^{1/\beta}$, $x > 0$. A equação $y = x^{1/\beta}$ tem a solução $x = y^{\beta}$ que nos dá $dx/dy = \beta y^{\beta - 1}$. Assim, de acordo com (18), Y tem a densidade g dada por

$$g(y) = \begin{cases} |\beta| \lambda y^{\beta - 1} e^{-\lambda y^{\beta}}, & y > 0, \\ 0, & y \leqslant 0. \end{cases}$$

Exemplo 8. Seja X uma variável aleatória contínua com densidade f e sejam a e b duas constantes, com $b \neq 0$. Então, de acordo com o Teorema 1, a densidade da variável aleatória $Y = a + bX$ é dada por

$$g(y) = \frac{1}{|b|} f\left(\frac{y - a}{b}\right), \quad -\infty < y < \infty. \tag{19}$$

Como uma ilustração desta fórmula, seja X a variável aleatória definida no Exemplo 1. No Exemplo 4 determinamos sua densidade f como sendo $f(x) = 2x/R^2$ para $0 < x < R$ e $f(x) = 0$ para outros valores de x. Considere a variável aleatória $Y = X/R$ e seja g sua densidade. Então, pela Fórmula (19) com $a = 0$ e $b = 1/R$,

$$g(y) = Rf(Ry) = 2y, \quad 0 < y < 1.$$

e $g(y) = 0$ para outros valores de y.

Se o leitor preferir, pode derivar fórmulas como as dos Exemplos 7 e 8 usando o método direto do Exemplo 6 em vez do Teorema 1.

Como vimos nos exemplos acima, podemos construir funções de densidade considerando funções de variáveis aleatórias. Existe outra maneira simples de construir funções de densidade. Seja g uma função não-negativa qualquer, tal que

$$0 < \int_{-\infty}^{\infty} g(x)\,dx < \infty.$$

Então sempre se pode normalizar g para obter uma função de densidade $f = c^{-1} g$, onde c é a constante

$$c = \int_{-\infty}^{\infty} g(x)\,dx.$$

Os exemplos seguintes ilustram este método.

Exemplo 9. Seja $g(x) = x(1-x)$ se $0 \leqslant x \leqslant 1$ e $g(x) = 0$ para outros valores de x. Então

$$c = \int_0^1 x(1-x)\,dx = \left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Bigg|_0^1 = \frac{1}{6}$$

e $f = c^{-1}\ g$ é dada por $f(x) = 6x\,(1-x)$ se $0 \leqslant x \leqslant 1$ e $f(x) = 0$ para outros valores de x. A função de distribuição correspondente é dada por $F(x) = 0$ para $x < 0, F(x) = 3x^2 - 2x^3$ para $0 \leqslant x \leqslant 1$ e $F(x) = 1$ para $x > 1$.

Exemplo 10. Seja $g(x) = 1/(1+x^2)$, $-\infty < x < \infty$. Sabemos de cálculo que a integral indefinida de $1/(1+x^2)$ é $\operatorname{arctg} x$. Assim

$$c = \int_{-\infty}^{\infty} \frac{dx}{1+x^2} = \operatorname{arctg} x \Bigg|_{-\infty}^{\infty} = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi.$$

Conseqüentemente $f = c^{-1}\ g$ é dada por

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty.$$

Esta densidade é conhecida como a *densidade de Cauchy*. A função de distribuição correspondente é dada por

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x, \qquad -\infty < x < \infty.$$

Como ilustração de uma variável aleatória com distribuição de Cauchy, temos o seguinte:

Exemplo 11. Seja X a tangente de um ângulo (medido em radianos) escolhido ao acaso de $(-\pi/2, \pi/2)$. Determine a distribuição de X.

Na resolução deste problema designaremos por Θ a variável aleatória que representa o ângulo escolhido, medido em radianos. Então $X = \operatorname{tg} \Theta$ e portanto (ver Figura 5) para $-\infty < x < \infty$.

$$\begin{aligned} P(X \leq x) &= P(\operatorname{tg} \Theta \leq x) \\ &= P\left(-\frac{\pi}{2} < \Theta \leq \operatorname{arctg} x\right) \\ &= \frac{1}{\pi}\left(\operatorname{arctg} x - \left(-\frac{\pi}{2}\right)\right) \\ &= \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x. \end{aligned}$$

Assim X tem a distribuição de Cauchy

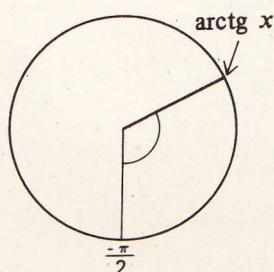



Figura 5

5.2.2. **DENSIDADES SIMÉTRICAS.** Encerraremos esta seção discutindo densidades simétricas e variáveis aleatórias simétricas. Diz-se que uma função de densidade f é simétrica se $f(x) = f(-x)$ para todo x. A densidade de Cauchy e a densidade uniforme em $(-a, a)$ são ambas simétricas. Diz-se que uma variável aleatória X é simétrica se X e $-X$ têm a mesma função de distribuição. O resultado do seguinte mostra que estes dois conceitos são intimamente relacionados.

Teorema 2. Seja X uma variável aleatória que possui uma densidade, então X tem uma densidade simétrica se, e somente se, X for uma variável aleatória simétrica.

Demonstração. Demonstraremos este resultado para variáveis aleatórias contínuas. A demonstração é similar para variáveis aleatórias discretas. Na nossa demonstração usaremos fato de que para qualquer função integrável f

$$\int_{-\infty}^{x} f(-y)\, dy = \int_{-x}^{\infty} f(y)\, dy, \qquad -\infty < x < \infty.$$

Suponha inicialmente que X tenha uma densidade simétrica f. Então

$$
\begin{aligned}
P(-X \leq x) &= P(X \geq -x) \\
&= \int_{-x}^{\infty} f(y)\,dy \\
&= \int_{-\infty}^{x} f(-y)\,dy \\
&= \int_{-\infty}^{x} f(y)\,dy \\
&= P(X \leq x),
\end{aligned}
$$

de modo que X e $-X$ têm a mesma função de distribuição.

Suponha ao contrário que X e $-X$ tenham uma densidade comum g. Defina f através de $f(x) = (g(x) + g(-x))/2$. Então f é claramente uma função de densidade simétrica. Além disso

$$
\begin{aligned}
\int_{-\infty}^{x} f(y)\,dy &= 1/2 \int_{-\infty}^{x} g(y)\,dy + 1/2 \int_{-\infty}^{x} g(-y)\,dy \\
&= 1/2 \int_{-\infty}^{x} g(y)\,dy + 1/2 \int_{-x}^{\infty} g(y)\,dy \\
&= 1/2[P(X \leq x)] + 1/2[P(-X \geq -x)] \\
&= P(X \leq x).
\end{aligned}
$$

Assim X possui a densidade simétrica f como desejado.

Se uma função de distribuição contínua F tem uma densidade simétrica f, então $F(0) = 1/2$. Os valores de F para x negativos podem ser calculados a partir dos valores de F para x positivos, pois

$$
\begin{aligned}
F(-x) &= \int_{-\infty}^{-x} f(y)\,dy \\
&= \int_{x}^{\infty} f(-y)\,dy \\
&= \int_{x}^{\infty} f(y)\,dy \\
&= \int_{-\infty}^{\infty} f(y)\,dy - \int_{-\infty}^{x} f(y)\,dy
\end{aligned}
$$

e portanto

(20) $$F(-x) = 1 - F(x), \qquad -\infty < x < \infty.$$

Por esta razão, quando se constrói tabelas de uma função de distribuição simétrica, geralmente só se apresenta valores não-negativos de x.

5.3. DENSIDADES NORMAL, EXPONENCIAL E GAMA

Discutiremos nesta seção três das mais importantes famílias de funções de densidade na teoria da probabilidade e estatística.

5.3.1 DENSIDADE NORMAL. Seja $g(x) = e^{-x^2/2}$, $-\infty < x < \infty$. Para normalizar g e transformá-la em uma densidade, precisamos avaliar a constante

$$c = \int_{-\infty}^{\infty} e^{-x^2/2}\, dx.$$

Não existe nenhuma fórmula simples para a integral indefinida de $e^{-x^2/2}$. A maneira mais fácil de avaliar c é através de um artifício especial em que escrevemos c como uma integral bidimensional e introduzimos as coordenadas polares. Para ser específico

$$\begin{aligned} c^2 &= \int_{-\infty}^{\infty} e^{-x^2/2}\, dx \int_{-\infty}^{\infty} e^{-y^2/2}\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy \\ &= \int_{0}^{\infty} \left(\int_{-\pi}^{\pi} e^{-r^2/2} r\, d\theta \right) dr \\ &= 2\pi \int_{0}^{\infty} r e^{-r^2/2}\, dr \\ &= -2\pi e^{-r^2/2} \Big|_{0}^{\infty} \\ &= 2\pi. \end{aligned}$$

Assim $c = \sqrt{2\pi}$ e a forma normalizada de g é dada por

$$f(x) = (2\pi)^{-1/2} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

Deixamos também registrada a fórmula

$$\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}. \tag{21}$$

A densidade que acabamos de derivar chama-se *densidade normal padrão* e representa-se geralmente por φ, de modo que

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty. \tag{22}$$

A densidade normal padrão é claramente simétrica. Representa-se a função de distribuição de φ por Φ. Não existe nenhuma fórmula simples para Φ de modo que ela deve ser avaliada numericamente. Existem rotinas de computador e tabelas

como a Tabela I no final deste livro para determinar Φ. Uma vez que Φ é simétrica, (20) é aplicável e

(23) $$\Phi(-x) = 1 - \Phi(x), \qquad -\infty < x < \infty.$$

Seja X uma variável aleatória com densidade normal padrão φ e seja $Y = \mu + \sigma X$, onde $\sigma > 0$. Então pela Fórmula (19) Y tem a densidade g dada por

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2}, \qquad -\infty < y < \infty.$$

Esta densidade chama-se densidade normal de média μ e variância σ^2, sendo representada por $n(\mu, \sigma^2)$ ou $n(y; \mu, \sigma^2)$, $-\infty < y < \infty$. Portanto

(24) $$n(y; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2} = \frac{1}{\sigma}\varphi\left(\frac{y-\mu}{\sigma}\right), \quad -\infty < y < \infty.$$

Como não definimos ainda momentos de variáveis aleatórias contínuas, podemos pensar temporariamente em μ e σ^2 como os parâmetros da família de densidades normais. A função de distribuição correspondente pode ser calculada em termos de Φ, pois

$$\begin{aligned} P(Y \leq y) &= P(\mu + \sigma X \leq y) \\ &= P\left(X \leq \frac{y-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{y-\mu}{\sigma}\right). \end{aligned}$$

Segue-se que se Y tiver a distribuição $n(\mu, \sigma^2)$ e $a \leqslant b$, então

(25) $$P(a \leqslant Y \leqslant b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Suponha por exemplo, que Y tenha a densidade $n(1, 4)$ e sejam $a = 0$ e $b = 3$. Da Tabela I obtemos que

$$\begin{aligned} P(0 \leqslant Y \leqslant 3) &= \Phi(1) - \Phi(-1/2) = \Phi(1) - (1 - \Phi(1/2)) \\ &= 0{,}8413 - 0{,}3085 \\ &= 0{,}5328. \end{aligned}$$

Se uma variável aleatória Y se distribui segundo $n(\mu, \sigma^2)$, então a variável aleatória $a + bY$, $b \neq 0$, se distribui segundo $n(a + b\mu, b^2\sigma^2)$. Isto é uma apli-

cação direta de (19). Alternativamente podemos escrever $Y = \mu + \sigma X$ com X tendo distribuição normal padrão. Então

$$a + bY = a + b(\mu + \sigma X) = (a + b\mu) + b\sigma X,$$

que se distribui segundo $n(a + b\mu, b^2 \sigma^2)$.

Variáveis aleatórias normalmente distribuídas ocorrem com bastante freqüência em aplicações práticas. A Lei de Maxwell em física estabelece que, sob condições apropriadas, os componentes da velocidade de uma molécula de gás se distribuirão segundo uma densidade normal $n(0, \sigma^2)$ em que σ^2 depende de certas quantidades físicas. Na maioria das aplicações entretanto, as variáveis aleatórias de interesse têm função de distribuição que é apenas aproximadamente normal. Por exemplo, constatou-se empiricamente que os erros de medida em experimentos físicos, a variabilidade da eficiência das linhas de produção e a variabilidade biológica (como a do peso e a da altura) têm distribuição aproximadamente normais. Constatou-se também, tanto empírica como teoricamente, que flutuações aleatórias resultantes da combinação de um grande número de causas não relacionadas, insignificantes individualmente, tendem a se distribuirem de forma aproximadamente normal. Os resultados teóricos neste sentido são conhecidos como "teoremas do limite central" e desenvolveram-se em um dos mais importantes tópicos de pesquisa em teoria da probabilidade. Um destes teoremas será discutido no Capítulo 7 e demonstrado no Capítulo 8. A importância das distribuições normais decorre também de suas atraentes propriedades teóricas. Um exemplo é a propriedade segundo a qual a soma de variáveis aleatórias independentes normalmente distribuídas também se distribui normalmente. Este fato será demonstrado no Capítulo 6. No volume II veremos que distribuições normais também desempenham um papel fundamental em estatística teórica e aplicada.

5.3.2. DENSIDADES EXPONENCIAIS. A densidade exponencial de parâmetro λ foi definida na Seção 5.2. Ela é dada por

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0, \\ 0, & x < 0, \end{cases} \tag{26}$$

A função de distribuição correspondente é

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geqslant 0, \\ 0 & x < 0. \end{cases} \tag{27}$$

Da discussão do Capítulo 1 e Exemplo 2 deste capítulo vemos que variáveis aleatórias, exponencialmente distribuídas, são úteis para o estudo de tempos até

a desintegração de partículas radioativas. Também são úteis para desenvolver modelos envolvendo muitos outros tempos de espera, tais como o tempo até a falha de um equipamento, o tempo necessário para completar uma tarefa ou o tempo que decorre até a chegada de um novo usuário. As variáveis aleatórias, exponencialmente distribuídas, têm também importância teórica, como se pode ver estudando os processos de Poisson (Capítulo 9) ou as cadeias de Markov no tempo contínuo (Volume III).

Uma importante propriedade de variáveis aleatórias exponencialmente distribuídas é que se X for uma tal variável, então

$$P(X > a + b) = P(X > a)P(X > b), \quad a \geq 0 \text{ e } b \geq 0. \tag{28}$$

(Esta fórmula é similar àquela obtida no Capítulo 3 para variáveis aleatórias geometricamente distribuídas.) Para ver que (28) é válida, seja λ o parâmetro da distribuição exponencial de X. Então, de (27) temos

$$\begin{aligned} P(X > a)P(X > b) &= e^{-\lambda a}e^{-\lambda b} \\ &= e^{-\lambda(a+b)} \\ &= P(X > a + b). \end{aligned}$$

Uma forma equivalente porém mais sugestiva de (28) é

$$P(X > a + b \mid X > a) = P(X > b), \quad a \geq 0 \text{ e } b \geq 0. \tag{29}$$

Pense em X como o tempo que decorre até a falha de um equipamento após sua instalação. Então (29) afirma que, sob a condição de que o equipamento não falhe até o tempo a, a probabilidade de falhar além de não ocorrer nas próximas b unidades de tempo, é igual à probabilidade incondicional do equipamento não falhar nas primeiras b unidades de tempo. Isto implica que o envelhecimento do equipamento não aumenta nem diminui sua probabilidade de falhar num dado intervalo de tempo.

O teorema seguinte mostra que (28) ou (29) caracteriza a família das distribuições exponenciais.

Teorema 3. Seja X uma variável aleatória tal que (28) se verifica. Então ou $P(X > 0) = 0$ ou X tem distribuição exponencial.

Demonstração. Se $P(X > 0) = 0$, então (28) se verifica trivialmente. Suponha que (28) se verifique e que $P(X > 0) \neq 0$. Então vemos de (28) com $a = b = 0$ que $P(X > 0) = 1$, de modo que X é uma variável aleatória positiva. Seja F a função de distribuição de X e defina G através de $G(x) = 1 - F(x) = P(X > x)$. Então G é uma função não crescente, contínua à direita, $G(0) = 1$, $G(+\infty) = 0$, e de acordo com (28)

$$G(a + b) = G(a)G(b), \quad a > 0 \text{ e } b > 0.$$

Segue-se que se $c > 0$ e m e n forem inteiros positivos, então

$$(30) \qquad G(nc) = (G(c))^n \quad \text{e} \quad G(c) = (G(c/m))^m.$$

Afirmamos a seguir que $0 < G(1) < 1$. Pois se $G(1) = 1$, então $G(n) = (G(1))^n = 1$, o que contradiz $G(+\infty) = 0$. Se $G(1) = 0$, então $G(1/m) = 0$ e pela continuidade à direita, $G(0) = 0$, outra contradição.

Uma vez que $0 < G(1) < 1$, podemos escrever $G(1) = e^{-\lambda}$ onde $0 < \lambda < \infty$. Segue-se de (30) se m for um inteiro positivo, então $G(1/m) = e^{-\lambda/m}$. Uma segunda aplicação de (30) mostra que se m e n forem inteiros positivos, então $G(n/m) = e^{-\lambda n/m}$. Em outras palavras, $G(y) = e^{-\lambda y}$ é válida para todo número racional positivo y. Pela continuidade à direita, segue-se que $G(y) = e^{-\lambda y}$ para $y \geq 0$. Isto implica que $F = 1 - G$ é a função de distribuição exponencial de parâmetro λ.

5.3.3. DENSIDADES GAMA. Antes de definirmos as densidades gama em geral, discutiremos um exemplo em que elas surgem naturalmente.

Exemplo 12. Seja X uma variável aleatória com densidade normal $n(0, \sigma^2)$. Obtenha a densidade da variável aleatória $Y = X^2$.

Para resolver este problema observamos inicialmente que a densidade de X é

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, \qquad -\infty < x < \infty.$$

De acordo com a fórmula (13), Y tem densidade g dada por $g(y) = 0$ para $y \leq 0$ e

$$g(y) = \frac{1}{2\sqrt{y}} (f(\sqrt{y}) + f(-\sqrt{y})), \qquad y > 0.$$

Isto implica que

$$(31) \qquad g(y) = \frac{1}{\sigma\sqrt{2\pi y}} e^{-y/2\sigma^2}, \qquad y > 0.$$

Para definir densidades gama em geral, consideramos inicialmente funções g da forma

$$g(x) = \begin{cases} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Aqui exigimos que $\alpha > 0$ e $\lambda > 0$ para que g seja integrável. A densidade em (31) corresponde ao caso especial $\alpha = 1/2$ e $\lambda = 1/2\sigma^2$. Na normalização de g para transformá-lo em densidade devemos avaliar

$$c = \int_0^\infty x^{\alpha-1} e^{-\lambda x}\, dx.$$

Fazendo a mudança de variável $y = \lambda x$, temos

$$c = \frac{1}{\lambda^\alpha} \int_0^\infty y^{\alpha-1} e^{-y}\, dy.$$

Não existe uma fórmula simples para a integral acima. Por isso ela é usada para definir uma função chamada função gama e é representada por Γ. Assim

$$c = \frac{1}{\lambda^\alpha} \Gamma(\alpha),$$

onde

$$(32) \qquad \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx, \qquad \alpha > 0.$$

A função normalizada chama-se densidade gama de parâmetros α e λ e é representada por $\Gamma(\alpha, \lambda)$ ou $\Gamma(x; \alpha, \lambda)$. Vemos que

$$(33) \qquad \Gamma(x; \alpha, \lambda) = \begin{cases} \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Deixamos também registrada a fórmula seguinte, que se mostrará útil no futuro:

$$(34) \qquad \int_0^\infty x^{\alpha-1} e^{-\lambda x}\, dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}.$$

As densidades exponenciais são casos especiais de densidades gama. Especificamente, a densidade de parâmetro λ é a mesma que a densidade $\Gamma(1, \lambda)$. Vimos também que a densidade dada por (31) é a densidade gama de parâmetros $\alpha = 1/2$ e $\lambda = 1/2\sigma^2$. Em outras palavras, *se* X *tiver a densidade* $n(0, \sigma^2)$, *então* X^2 *terá a densidade gama* $\Gamma(1/2, 1/2\sigma^2)$. Igualando (31) a (33) com $\alpha = 1/2$ e $\lambda = 1/2\sigma^2$, obtemos a útil relação

$$(35) \qquad \Gamma(1/2) = \sqrt{\pi}.$$

Uma importante propriedade da função gama é que

$$(36) \qquad \Gamma(\alpha + 1) = \alpha\Gamma(\alpha), \qquad \alpha > 0.$$

Esta fórmula decorre de (32) por uma simples aplicação de integração por partes. Para sermos específicos

$$\begin{aligned}\Gamma(\alpha + 1) &= \int_0^\infty x^\alpha e^{-x}\,dx \\ &= -x^\alpha e^{-x}\Big|_0^\infty + \int_0^\infty \alpha x^{\alpha-1} e^{-x}\,dx \\ &= \alpha\Gamma(\alpha).\end{aligned}$$

Como $\Gamma(1) = 1$, segue-se facilmente de (36) que se n for um positivo inteiro,

$$\Gamma(n) = (n-1)!. \tag{37}$$

Segue-se também de (35) e (36), após algumas simplificações, que se n for um positivo inteiro ímpar, então

$$\Gamma\left(\frac{n}{2}\right) = \frac{\sqrt{\pi}(n-1)!}{2^{n-1}\left(\frac{n-1}{2}\right)!}. \tag{38}$$

Não existe uma fórmula simples para a função de distribuição correspondente a $\Gamma(\alpha, \lambda)$ exceto quando $\alpha = m$ é um inteiro positivo. Neste caso podemos integrar por partes e obter para $x > 0$

$$\begin{aligned}\int_0^x \frac{\lambda^m y^{m-1} e^{-\lambda y}}{(m-1)!}\,dy &= \frac{-(\lambda y)^{m-1} e^{-\lambda y}}{(m-1)!}\Big|_0^x + \int_0^x \frac{\lambda^{m-1} y^{m-2} e^{-\lambda y}}{(m-2)!}\,dy \\ &= \int_0^x \frac{\lambda^{m-1} y^{m-2} e^{-\lambda y}}{(m-2)!}\,dy - \frac{(\lambda x)^{m-1} e^{-\lambda x}}{(m-1)!},\end{aligned}$$

desde que $m \geqslant 2$. Se integrarmos por partes $m - 1$ vezes desta maneira e observarmos que

$$\int_0^x \lambda e^{-\lambda y}\,dy = 1 - e^{-\lambda x},$$

obtemos a fórmula

$$\int_0^x \frac{\lambda^m y^{m-1} e^{-\lambda y}}{(m-1)!}\,dy = 1 - \sum_{k=0}^{m-1} \frac{(\lambda x)^k e^{-\lambda x}}{k!}, \qquad x > 0. \tag{39}$$

Esta fórmula reflete uma interessante relação entre uma variável aleatória X com densidade gama $\Gamma(m, \lambda)$ e uma variável aleatória Y com distribuição de Poisson de parâmetro λx. Especificamente, (39) estabelece que

$$P(X \leqslant x) = P(Y \geqslant m). \tag{40}$$

Como veremos no Capítulo 9, esta relação é relevante à teoria dos processos de Poisson.

O comportamento qualitativo da densidade gama, ilustrado na Figura 6, é facilmente obtido pelos métodos de cálculo.

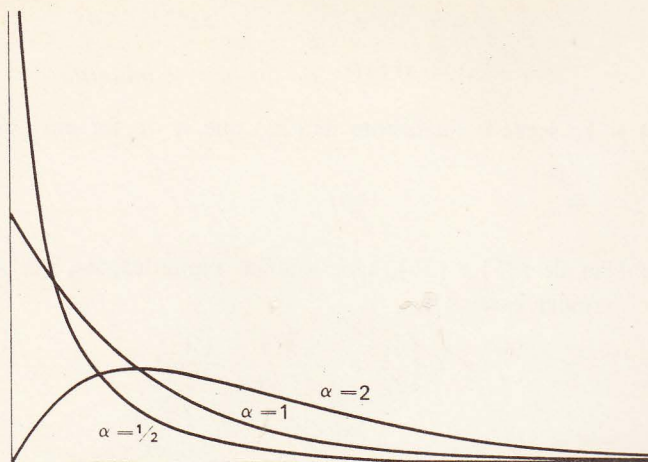



Figura 6. A Densidade Gama

Uma importante propriedade das densidades gama é que se X e Y forem variáveis aleatórias independentes tendo densidades $\Gamma(\alpha_1, \lambda)$ e $\Gamma(\alpha_2, \lambda)$ respectivamente, então $X + Y$ terá a densidade gama $\Gamma(\alpha_1 + \alpha_2, \lambda)$. Este resultado será demonstrado no Capítulo 6. Esta e outras propriedades das densidades gama tornam sua manipulação bastante simples. Existem muitas situações práticas em que não se conhece a densidade de uma variável aleatória X. É possível que se saiba que X é uma variável aleatória positiva cuja densidade pode ser aproximada razoavelmente por uma densidade gama de parâmetros apropriados. Em tais casos, resolvendo um problema envolvendo X, sob a hipótese que X tenha uma densidade gama, proporcionará uma aproximação ou pelo menos uma compreensão melhor da situação real desconhecida.

5.4. FUNÇÕES INVERSAS DE DISTRIBUIÇÃO

Importantes aplicações das fórmulas de mudança de variáveis da Seção 5.2.1. podem ser obtidas relacionando a função φ à função de distribuição F.

Seja X uma variável aleatória contínua tendo função de distribuição F e função de densidade f. Aplicaremos a fórmula de mudança de variável à função $\varphi = F$. Se $y = F(x)$, então $dy/dx = F'(x) = f(x)$ e portanto $dx/dy = 1/f(x)$. Assim de acordo com (18), a variável aleatória $Y = F(X)$ tem densidade g onde

$$g(y) = \frac{f(x)}{f(x)} = 1, \qquad 0 < y < 1,$$

e $g(y) = 0$ para outros valores de Y. Em outras palavras, a variável aleatória $Y = F(X)$ distribui-se uniformemente em $(0,1)$. Este resultado é válido mesmo que a função $\varphi = F$ não satisfaça todas as hipóteses do Teorema 1. Usando um argumento direto, pode-se mostrar que *se X é uma variável aleatória contínua com função de distribuição F, então $F(X)$ distribui-se uniformemente em* $(0, 1)$. (Se F for descontínua em algum ponto x_0, então $P(X = x_0) > 0$, de modo que $P(F(X) = F(X_0)) > 0$ e $F(X)$ não poderia distribuir-se uniformemente em $(0, 1)$.)

Pode-se também prosseguir em outra direção. Seja F uma função de distribuição contínua que é estritamente crescente em algum intervalo I e tal que, $F = 0$ à esquerda de I se I for limitado inferiormente e, $F = 1$ à direita de I se I for limitado superiormente. Então para $0 < y < 1$, pelo teorema do valor médio de cálculo, existe um único valor de x tal que $y = F(x)$. Assim $F^{-1}(y)$, $0 < y < 1$, é bem definida. Sob estas condições, *se Y for uma variável aleatória uniformemente distribuída em* $(0,1)$ *então a variável aleatória $F^{-1}(Y)$ tem F como sua função de distribuição.*

Pode-se usar dois exemplos da Seção 5.2.1 para ilustrar o resultado acima. No Exemplo 6 obtivemos variáveis aleatórias exponencialmente distribuídas como transformadas de uma variável aleatória uniformemente distribuída. O leitor poderia checar e veria que estas transformações poderiam ser obtidas pelo método do parágrafo acima. No Exemplo 11 mostramos que se Θ for uniformemente distribuída em $(-\pi/2, \pi/2)$, então $\operatorname{tg} \Theta$ terá a distribuição de Cauchy. Suponha que Y se distribua uniformemente em $(0, 1)$. Então $\Theta = \pi Y - \pi/2$ distribui-se uniformemente em $(-\pi/2, \pi/2)$, de modo que

$$X = \operatorname{tg} \Theta = \operatorname{tg} \left(\pi Y - \frac{\pi}{2}\right)$$

terá a distribuição de Cauchy. Este é exatamente o que obteríamos usando o resultado do parágrafo anterior. De acordo com o Exemplo 10, a função de distribuição de Cauchy é dada por

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x, \qquad -\infty < x < \infty,$$

e a equação $y = F(x)$, ou

$$y = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x,$$

tem a solução

$$x = F^{-1}(y) = \operatorname{tg} \left(\pi y - \frac{\pi}{2}\right).$$

Para alguns propósitos é desejável gerar uma variável aleatória X tendo uma função de distribuição pré-estabelecida F. Uma maneira de fazê-lo é gerar inicialmente uma variável aleatória uniformemente distribuída Y e então fazer $X = F^{-1}(Y)$. Este método é especialmente útil em computadores digitais já que existem métodos bastante satisfatórios para gerar (o que se comporta como) variáveis aleatórias uniformemente distribuídas em tais computadores. Suponha por exemplo que desejamos uma rotina para gerar uma variável aleatória X com densidade normal padrão $n(0,1)$. Usaríamos uma sub-rotina para gerar uma variável aleatória Y uniformemente distribuída em $(0,1)$ e uma sub-rotina para computar a função numérica Φ^{-1}, e então determinar $X = \Phi^{-1}(Y)$. Para gerar uma variável aleatória X tendo a densidade normal $n(\mu, \sigma^2)$ faríamos $X = \mu + \sigma\Phi^{-1}(Y)$.

As funções inversas de distribuição são úteis para outros propósitos. Suponha por exemplo que X tenha a densidade normal $n(\mu, \sigma^2)$ e lembre-se da Seção 5.3.1. que

$$P(X \leqslant b) = \Phi\left(\frac{b - \mu}{\sigma}\right)$$

Suponha que desejamos escolher b tal que $P(X \leqslant b) = 0{,}9$. Precisamos resolver em b a equação

$$\Phi\left(\frac{b - \mu}{\sigma}\right) = 0{,}9.$$

A solução é dada por

$$\left(\frac{b - \mu}{\sigma}\right) = \Phi^{-1}(0{,}9)$$

ou

$$b = \mu + \sigma\Phi^{-1}(0{,}9).$$

Da Tabela I vemos que $\Phi^{-1}(0{,}9) = 1{,}28$. Assim $b = \mu + 1{,}28\sigma$ e

$$P(X \leqslant \mu + 1{,}28\sigma) = 0{,}9.$$

Em estatística aplicada o número $b = \mu + 1{,}28\sigma$ chama-se decil superior da distribuição $n(\mu, \sigma^2)$.

Seja F uma função de distribuição qualquer que satisfaz as condições para que $F^{-1}(y)$, $0 < y < 1$, seja bem definida, conforme a discussão acima. Então $m = F^{-1}(1/2)$ chama-se *mediana* de F, $F^{-1}(3/4)$ e $F^{-1}(1/4)$ chama-se quartis superior e inferior de F, $F^{-1}(0{,}9)$ chama-se decil superior e $F^{-1}(k/100)$ chama-se

k-percentil superior. Estas definições podem ser modificadas para aplicá-las à função arbitrária de distribuição, e em particular as funções discretas de distribuição.

Se X tiver uma densidade simétrica então X terá claramente a mediana $m = 0$. Para um exemplo mais interessante, seja X uma variável aleatória exponencialmente distribuída com parâmetro λ. Então sua mediana é dada por $1 - e^{-\lambda m} = 1/2$, que tem a solução $m = \lambda^{-1} \log 2$. Suponha que X represente o tempo para uma partícula radioativa se desintegrar. Então se tivermos um número bastante grande de tais partículas, esperaremos que a metade das partículas tenham se desintegrado até o tempo m. Em física este tempo chama-se *meia-vida* da partícula. Se observarmos a meia-vida m poderemos usá-la para determinar a taxa de desintegração $\lambda = m^{-1} \log 2$.

Para uma aplicação final das funções inversas de distribuição, seja X uma variável aleatória com densidade normal $n(\mu, \sigma^2)$ e suponha que desejamos determinar $a > 0$ tal que $P(\mu - a \leqslant X \leqslant \mu + a) = 0,9$. Então de acordo com (25) precisamos resolver em a a equação

$$\Phi\left(\frac{a}{\sigma}\right) - \Phi\left(-\frac{a}{\sigma}\right) = 0,9.$$

Como $\Phi(-x) = 1 - \Phi(x)$ para todo x, temos

$$2\Phi\left(\frac{a}{\sigma}\right) - 1 = 0,9$$

e portanto $a = \sigma\Phi^{-1}(0,95)$. Da Tabela I vemos que $\Phi^{-1}(0,95) = 1,645$. Em outras palavras,

$$P(\mu - 1,645\sigma \leqslant X \leqslant \mu + 1,645\sigma) = 0,9.$$

Usando a mesma técnica obtemos

$$P(\mu - 0,675\sigma \leqslant X \leqslant \mu + 0,675\sigma) = 0,5$$

ou equivalente,

$$P(|X - \mu| \leqslant 0,675\sigma) = 0,5$$

Isto indica que se X tiver a densidade normal $n(\mu, \sigma^2)$, então X deferirá de μ por menos de $0,675\sigma$ com probabilidade um meio e por mais de $0,675\sigma$ com probabilidade um meio. Se pensarmos em μ como uma quantidade física real e X como uma medida de μ, então $|X - \mu|$ representa o erro de medida. Por esta razão $0,675\sigma$ é conhecido como o *erro provável*.

Exercícios

1. Seja X uma variável aleatória tal que $P(|X-1|=2)=0$. Expresse $P(|X-1| \geqslant 2)$ em termos da função de distribuição F_X.
2. Considere um ponto escolhido aleatoriamente no interior de um disco de raio R no plano. Seja X o quadrado da distância do centro do disco ao ponto escolhido. Obtenha a função de distribuição de X.
3. Considere um ponto escolhido uniformemente numa bola sólida de raio R no espaço tridimensional. Seja X a distância do centro da bola ao ponto escolhido. Obtenha a função de distribuição de X.
4. Considere um ponto escolhido uniformemente no intervalo $[0, a]$. Seja X a distância da origem ao ponto escolhido. Obtenha a função de distribuição de X.
5. Considere um ponto escolhido uniformemente no interior de um triângulo de base l e altura h. Seja X a distância da base ao ponto escolhido. Obtenha a função de distribuição de X.
6. Considere um triângulo eqüilátero de lado s. Seja um ponto escolhido uniformemente sobre um lado do triângulo. Seja X a distância entre o ponto escolhido e o vértice oposto. Obtenha a função de distribuição de X.
7. Seja o ponto (u, v) escolhido uniformemente no quadrado $0 \leqslant u \leqslant 1$, $0 \leqslant v \leqslant 1$. Seja X a variável aleatória que associa o número $u+v$ ao ponto (u, v). Obtenha a função de distribuição de X.
8. Seja F a função de distribuição dada pela fórmula (3). Obtenha um número m tal que $F(m)=1/2$.
9. Seja X o tempo até a desintegração de alguma partícula radioativa e suponha que a função de distribuição de X é dada pela fórmula (3). Suponha que λ seja tal que $P(X \geqslant 0{,}01)=1/2$. Obtenha um número t tal que $P(X \geqslant t)=0{,}9$.
10. Seja X a variável aleatória do Exercício 4. Obtenha a função de distribuição de $Y=\text{Min}\,(X, a/2)$.
11. Seja X uma variável aleatória cuja função de distribuição F é dada por

$$F(x)=\begin{cases} 0, & x<0, \\ \dfrac{x}{3}, & 0 \leq x<1, \\ \dfrac{x}{2}, & 1 \leq x<2, \\ 1, & x \geq 2. \end{cases}$$

Determine:

(a) $P(1/2 \leqslant X \leqslant 3/2)$;
(b) $P(1/2 \leqslant X \leqslant 1)$;
(c) $P(1/2 \leqslant X < 1)$;
(d) $P(1 \leqslant X \leqslant 3/2)$;
(e) $P(1 < X < 2)$.

12. Se a função de distribuição de X fosse definida de uma das formas abaixo, descreva como as propriedades (i)–(iv) da Seção 5.1.1 teriam de ser modificadas em cada caso:

(a) $F(x) = P(X < x)$;
(b) $F(x) = P(X > x)$;
(c) $F(x) = P(X \geqslant x)$.

13. Escolhe-se aleatoriamente um ponto em $(-10, 10)$. Seja X uma variável aleatória definida de tal forma que X represente a coordenada do ponto se o mesmo estiver em $[-5, 5]$, $X = -5$ se o ponto estiver em $(-10, -5)$ e $X = 5$ se o ponto estiver em $(5, 10)$. Obtenha a função de distribuição de X.

14. Seja X uma variável aleatória contínua tendo a densidade de f dada por

$$f(x) = (1/2)e^{-|x|}, \qquad -\infty < x < \infty.$$

Obtenha $P(1 \leqslant |X| \leqslant 2)$.

15. Seja F a função de distribuição definida por

$$F(x) = \frac{1}{2} + \frac{x}{2(|x| + 1)}, \qquad -\infty < x < \infty.$$

Obtenha uma densidade f para F. Para que valores de x teremos $F'(x) = f(x)$?

16. Obtenha uma função de densidade para a variável aleatória do Exercício 3.

17. Obtenha uma função de densidade para a variável aleatória do Exercício 7.

18. Seja X uma variável aleatória contínua de densidade f. Obtenha uma fórmula para a densidade de $Y = |X|$.

19. Sejam X e $Y = X^2$ duas variáveis aleatórias contínuas positivas tendo densidades f e g, respectivamente. Obtenha f em termos de g e g em termos de f.

20. Suponha que X se distribui uniformemente em (0,1). Obtenha a densidade de $Y = X^{1/\beta}$, onde $\beta \neq 0$.

21. Seja X uma variável aleatória contínua positiva de densidade f. Obtenha uma fórmula para a densidade de $Y = 1/(X + 1)$.

22. Seja X uma variável aleatória, g uma função de densidade em relação à integração e φ uma função diferenciável estritamente crescente em $(-\infty, \infty)$. Suponha que

$$P(X \leq x) = \int_{-\infty}^{\varphi(x)} g(z)\, dz, \qquad -\infty < x < \infty.$$

Mostre que a variável aleatória $Y = \varphi(X)$ tem densidade g.

23. Seja X uma variável aleatória com distribuição uniforme em (a, b). Obtenha uma função linear φ tal que $Y = \varphi(X)$ tenha distribuição uniforme em $(0, 1)$.

24. Suponha que X tenha uma densidade exponencial de parâmetro λ. Obtenha a densidade de $Y = cX$, onde $c > 0$.

25. Seja $g(x) = x(1 - x)^2$, $0 \leqslant x \leqslant 1$, e $g(x) = 0$ para outros valores de x. Como se deve normalizar g para transformá-la em uma densidade?

26. Suponha que X tenha a densidade de Cauchy. Obtenha a densidade de $Y = a + bX$, $b \neq 0$.

27. Seja X o seno de um ângulo escolhido aleatoriamente em $(-\pi/2, \pi/2)$. Obtenha a densidade e a função de distribuição de X.

28. Seja X uma variável aleatória contínua com densidade simétrica f e tal que X^2 tenha uma densidade exponencial de parâmetro λ. Obtenha f.

29. Seja X uma variável aleatória contínua tendo uma função de distribuição F e uma função de densidade f. Diz-se então que f é simétrica em relação a a se $f(a + x) = f(a - x)$, $-\infty < x < \infty$. Obtenha as condições equivalentes em termos de variável aleatória X e em termos de função de distribuição F.

30. Define-se a função de erro por

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2}\, dy, \qquad -\infty < x < \infty.$$

Expresse Φ em termos de função de erro.

31. Suponha que X tem a densidade normal $n(0, \sigma^2)$. Obtenha a densidade de $Y = |X|$.

32. Suponha que X tem a densidade normal $n(\mu, \sigma^2)$. Obtenha a densidade de $Y = e^X$. Esta densidade chama-se *densidade log normal.*

33. Suponha que X se distribui normalmente com parâmetros μ e σ^2. Obtenha $P(|X - \mu| \leqslant \sigma)$.

34. Suponha que X se distribui normalmente com parâmetros μ e σ^2. Determine os números a e b tais que $a + bX$ tenha a densidade normal padrão.

35. Suponha que X se distribui normalmente com parâmetros $\mu = 0$ e $\sigma^2 = 4$. Seja Y uma variável aleatória inteira definida em termos de X por $Y = m$ se $m - 1/2 \leqslant X < m + 1/2$, onde m é um número inteiro tal que $-5 \leqslant m \leqslant 5$, $Y = -6$ se $X < -5{,}5$ e $Y = 6$ se $X \geqslant 5{,}5$. Obtenha f_Y e faça um gráfico desta densidade.

36. Suponha que o peso de uma pessoa selecionada ao acaso de uma certa população distribui-se normalmente com parâmetros μ e σ. Suponha também que $P(X \leqslant 160) = 1/2$ e $P(X \leqslant 140) = 1/4$. Obtenha μ e σ e determine

$P(X \geqslant 200)$. De todas as pessoas pesando no mínimo 200 libras, que percentagem pesará mais de 220 libras?

37. Seja t_p um número tal que $\Phi(t_p) = p$, $0 < p < 1$. Suponha que X tem a densidade normal $n(\mu, \sigma^2)$. Mostre que para $0 < p_1 < p_2 < 1$,

$$P(\mu + t_{p_1}\sigma \leqslant X \leqslant \mu + t_{p_2}\sigma) = p_2 - p_1.$$

38. Suponha que um número bastante grande de partículas radioativas idênticas tenha tempo de desintegração que se distribua exponencialmente com um certo parâmetro λ. Se a metade das partículas se desintegram no primeiro segundo, quanto tempo levará para 75% das partículas se desintegrarem?

39. Suponha que X se distribui exponencialmente com parâmetro λ. Seja Y a variável aleatória inteira definida em termos de X por $Y = m$ se $m \leqslant X < m + 1$, onde m é um número inteiro não-negativo. Como se distribui Y?

40. Seja T uma variável aleatória contínua positiva que representa o tempo de falha de um certo sistema, seja F a função de distribuição de T e suponha que $F(t) < 1$ para $0 < t < \infty$. Então podemos escrever $F(t) = 1 - e^{-G(t)}$, $t > 0$. Suponha que $G'(t) = g(t)$ existe para $t > 0$.

(a) Mostre que T tem densidade f dada por

$$\frac{f(t)}{1 - F(t)} = g(t), \qquad 0 < t < \infty.$$

A função g é conhecida como "taxa de falha", pois heuristicamente

$$P(t \leqslant T \leqslant t + dt \mid T > t) = \frac{f(t)\,dt}{1 - F(t)} = g(t)\,dt.$$

(b) Mostre que para $s > 0$ e $t > 0$,

$$P(T > t + s \mid T > t) = e^{-\int_t^{t+s} g(u)\,du}.$$

(c) Mostre que o sistema melhora com a idade (isto é, para um s fixo a expressão em (b) cresce com t) se g for uma função decrescente, e o sistema deteriora com a idade se g for uma função crescente.

(d) Mostre que

$$\int_0^\infty g(u)\,du = \infty.$$

(e) Como se comporta g se T tem distribuição exponencial?

(f) Se $G(t) = \lambda t^\alpha$, $t > 0$, para que valores de α o sistema melhora, deteriora e não se altera com a idade?

41. Suponha que X tem a densidade gama $\Gamma(\alpha, \lambda)$. Obtenha a densidade de $Y = cX$, onde $c > 0$.

42. Mostre que se $\alpha > 1$, a densidade gama tem um máximo em $(\alpha - 1)/\lambda$.

43. Suponha que X tem a densidade gama $\Gamma(\alpha, \lambda)$. Obtenha a densidade de $Y = \sqrt{X}$.

44. Suponha que Y se distribui uniformemente em $(0,1)$. Obtenha uma função φ tal que $X = \varphi(Y)$ tenha a densidade f dada por $f(x) = 2x$, $0 \leqslant x \leqslant 1$ e $f(x) = 0$ para outros valores de x.

45. Suponha que Y se distribui uniformemente em $(0,1)$. Obtenha uma função φ tal que $\varphi(Y)$ tenha a densidade gama $\Gamma(1/2, 1/2)$. *Sugestão:* use o Exemplo 12.

46. Determine $\Phi^{-1}(t)$ para $t = 0{,}1; 0{,}2; \ldots ; 0{,}9$ e use estes valores para construir um gráfico de Φ^{-1}.

47. Suponha que X tem a densidade normal $n(\mu, \sigma^2)$. Obtenha o quartil superior de X.

48. Suponha que X tem a densidade de Cauchy. Determine o quartil superior de X.

49. Suponha que X tem a densidade normal com parâmetros μ e $\sigma^2 = 0{,}25$. Determine uma constante c tal que

$$P(|X - \mu| \leqslant c) = 0{,}9.$$

50. Seja X uma variável aleatória inteira com função de distribuição F, e suponha que Y se distribui uniformemente em $(0,1)$. Defina a variável aleatória inteira Z em termos de Y por

$$Z = m \quad \text{se} \quad F(m-1) < Y \leqslant F(m),$$

Para qualquer inteiro m. Mostre que Z tem a mesma densidade de X.

6

VARIÁVEIS ALEATÓRIAS COM DISTRIBUIÇÃO CONJUNTA

Nas primeiras três seções deste capítulo discutiremos um par de variáveis aleatórias contínuas X e Y e algumas de suas propriedades. Nas quatro seções restantes consideraremos as extensões de duas para n variáveis aleatórias $X_1, X_2, \ldots, X_n$. A discussão de estatísticas de ordem da Seção 6.5 é opcional e não será necessária nos capítulos subseqüentes deste livro. A Seção 6.6 é essencialmente um resumo dos resultados de distribuições amostrais que são úteis na estatística e serão necessários no Volume II. O material coberto na Seção 6.7 será usado somente na demonstração do Teorema 1 do Capítulo 9 e Teorema 1 do Capítulo 5 do Volume II.

6.1. PROPRIEDADES DAS DISTRIBUIÇÕES BIDIMENSIONAIS

Sejam X e Y duas variáveis aleatórias definidas no mesmo espaço de probabilidade. Define-se a sua *função de distribuição conjunta* F por

$$F(x, y) = P(X \leqslant x, Y \leqslant y), \qquad -\infty < x, y < \infty.$$

Para ver se F é bem definida, observe que $\{\omega / X(\omega) \leqslant x\}$ e $\{\omega / Y(\omega) \leqslant y\}$ ambos são eventos já que X e Y são variáveis aleatórias. A sua união $\{\omega / X(\omega) \leqslant x$ e $Y(\omega) \leqslant y\}$ é também um evento, e portanto sua probabilidade é bem definida.

Pode-se usar a função de distribuição conjunta para calcular a probabilidade de que o par (X, Y) esteja em um retângulo no plano. Considere o retângulo

$$R = \{(x, y) \mid a < x \leqslant b, c < y \leqslant d\},$$

onde $a \leqslant b$ e $c \leqslant d$. Então

$$(1) \qquad \begin{aligned} P((X, Y) \in R) &= P(a < X \leqslant b, c < Y \leqslant d) \\ &= F(b, d) - F(a, d) - F(b, c) + F(a, c). \end{aligned}$$

Para verificar que (1) é verdadeiro, observe que

$$P(a < X \leqslant b, Y \leqslant d) = P(X \leqslant b, Y \leqslant d) - P(X \leqslant a, Y \leqslant d)$$
$$= F(b, d) - F(a, d).$$

Analogamente

$$P(a < X \leqslant b, Y \leqslant c) = F(b, c) - F(a, c).$$

Assim

$$P(a < X \leq b, c < Y \leq d)$$
$$= P(a < X \leq b, Y \leq d) - P(a < X \leq b, Y \leq c)$$
$$= (F(b, d) - F(a, d)) - (F(b, c) - F(a, c))$$

e portanto (1) é verdadeiro como desejado.

As funções de distribuição unidimensional F_X e F_Y definidas por

$$F_X(x) = P(X \leqslant x) \quad \text{e} \quad F_Y(y) = P(Y \leqslant y)$$

são chamadas de *funções de distribuição marginal* de X e Y. Elas se relacionam com a função de distribuição conjunta F através de

$$F_X(x) = F(x. \infty) = \lim_{y \to \infty} F(x, y)$$

$$F_Y(y) = F(\infty, y) = \lim_{x \to \infty} F(x, y).$$

Se existir uma função não-negativa f tal que

$$(2) \qquad F(x, y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f(u, v)\, dv \right) du, \qquad -\infty < x, y < \infty,$$

então f é chamada de função de densidade conjunta (em relação a integração) da função de densidade conjunta F ou do par de variáveis aleatórias X e Y. A menos que se especifique em contrário, usaremos ao longo de todo este capítulo o termo funções de densidade para fazer referência a funções de densidade em relação à integração em vez de funções de densidade discreta.

Se F tiver densidade f, podemos reescrever a Equação (1) em termos de f, obtendo

$$(3) \qquad P(a < X \leq b, c < Y \leq d) = \int_a^b \left(\int_c^d f(x, y)\, dy \right) dx.$$

Usando as propriedades da integração e a definição de espaços de probabilidade, pode-se mostrar que a relação

$$(4) \qquad P((X, Y) \in A) = \iint_A f(x, y)\, dx\, dy$$

se verifica para subconjuntos A no plano do tipo considerado em cálculo. Fazendo A representar todo o plano, obtemos de (4) que

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1. \tag{5}$$

Obtemos também de (4) que

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x}\left(\int_{-\infty}^{\infty} f(u, y)\, dy\right) du$$

de modo que F_X tem densidade marginal f_X dada por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

que satisfaz

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du.$$

Analogamente F_Y tem densidade marginal f_Y dada por

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

Como no caso unidimensional, f não é unicamente definida por (2). Podemos alterar f em um número finito de pontos ou mesmo sobre um número finito de curvas suaves no plano sem afetar a integral de f sobre subconjuntos do plano. Novamente como no caso unidimensional, F determina f nos pontos de continuidade de f. Pode-se obter este fato de (3).

Diferenciando (2) e aplicando as regras de cálculo obtemos

$$\frac{\partial}{\partial y} F(x, y) = \int_{-\infty}^{x}\left(\frac{\partial}{\partial y}\int_{-\infty}^{y} f(u, v)\, dv\right) du$$

$$= \int_{-\infty}^{x} f(u, y)\, du$$

e

$$\frac{\partial^2}{\partial x\, \partial y} F(x, y) = f(x, y). \tag{6}$$

Sob algumas condições brandas adicionais podemos justificar estas operações e mostrar que (6) é válido nos pontos de continuidade de f. Em casos específicos, em vez de verificar se as etapas que conduzem a (6) são válidas, é geralmente mais simples mostrar que a função f obtida de (6) satisfaz (2).

Exemplo 1. Ilustraremos as definições e fórmulas acima reconsiderando o Exemplo 1 do Capítulo 5. Lembre-se que naquele exemplo escolhemos uniformemente um ponto de um disco de raio R. Suponha que os pontos do plano são determinados por suas coordenadas cartesianas (x, y). Então o disco pode ser representado como

$$\{(x, y) \mid x^2 + y^2 \leqslant R^2\}.$$

Sejam X e Y as variáveis aleatórias que representam as coordenadas aleatórias do ponto escolhido. Correspondendo à hipótese de uniformidade, supomos que X e Y tem densidade conjunta f dada por

$$(7) \qquad f(x, y) = \begin{cases} \dfrac{1}{\pi R^2}, & x^2 + y^2 \leq R^2, \\ 0, & \text{outros valores de } x \text{ e } y. \end{cases}$$

Então para qualquer subconjunto A do disco (digamos do tipo considerado em cálculo),

$$P((X, Y) \in A) = \iint_A f(x, y)\, dx\, dy$$

$$= \frac{\text{área de } A}{\pi R^2},$$

que concorda com nossa hipótese de uniformidade. A densidade marginal f_X é dada por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{\pi R^2}\, dy = \frac{2\sqrt{R^2 - x^2}}{\pi R^2}$$

para $-R < x < R$ e $f_X(x) = 0$ para outros valores de x. A densidade marginal $f_Y(y)$ é dada pela mesma fórmula com y substituindo x.

Diz-se que as variáveis X e Y são *variáveis aleatórias independentes* se $a \leqslant b$ e $c \leqslant d$, então

$$(8) \qquad P(a < X \leqslant b, c < Y \leqslant d) = P(a < X \leqslant b)P(c < Y \leqslant d).$$

sempre que $a = c = -\infty$, $b = x$, e $d = y$, segue-se que se X e Y forem independentes, então

$$(9) \qquad F(x, y) = F_X(x)F_Y(y), \qquad -\infty < x, y < \infty.$$

Reciprocamente (9) implica que X e Y são independentes. Pois se (9) for verdadeiro, então de acordo com (1) o primeiro membro de (8) é

$$\begin{aligned} F(b, d) - F(a, d) - F(b, c) + F(a, c) & \\ &= F_X(b)F_Y(d) - F_X(a)F_Y(d) - F_X(b)F_Y(c) + F_X(a)F_Y(c) \\ &= (F_X(b) - F_X(a))(F_Y(d) - F_Y(c)) \\ &= P(a < X \leq b)P(c < Y \leq d). \end{aligned}$$

Pode-se mostrar de uma forma mais geral que se X e Y forem independentes e A e B forem uniões de um número finito ou infinito enumerável de intervalos, então

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

ou em outras palavras os eventos

$$\{\omega \mid X(\omega) \in A\} \quad \text{e} \quad \{\omega \mid X(\omega) \in B\}$$

serão eventos independentes.

Sejam X e Y variáveis aleatórias tendo densidades marginais f_X e f_Y. Então X e Y são independentes se, e somente se, a função f definida por

$$f(x,y) = f_X(x) f_Y(y), \qquad -\infty < x, y < \infty,$$

for uma densidade conjunta para X e Y. Isto é uma conseqüência da definição de independência e da fórmula

$$F_X(x)F_Y(y) = \int_{-\infty}^{x} \left(\int_{-\infty}^{y} f_X(u)f_Y(v)\, dv \right) du.$$

Como uma ilustração de variáveis aleatórias dependentes, sejam X e Y as variáveis definidas no Exemplo 1. Então para $-R < x < R$ e $-R < y < R$,

$$(10) \qquad f_X(x)f_Y(y) = \frac{4\sqrt{R^2 - x^2}\,\sqrt{R^2 - y^2}}{\pi^2 R^4},$$

que não concorda com a densidade conjunta dessas variáveis aleatórias no ponto $x = 0$, $y = 0$. Como $(0,0)$ é um ponto de continuidade das funções definidas por (7) e (10), segue-se que X e Y são variáveis aleatórias dependentes. Isto está de acordo com a nossa noção intuitiva de dependência, pois quando X está próximo de R, Y deve estar próximo de zero, e assim informação sobre X nos dá informação sobre Y.

Pode-se definir diretamente as funções de densidade, como vimos em outros contextos. Uma *função de densidade bidimensional* f é uma função não-negativa em R^2 tal que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1.$$

Correspondendo a qualquer função de densidade bidimensional existem um espaço de probabilidade e um par de variáveis aleatórias X e Y definidas neste espaço cuja densidade conjunta é f.

A maneira mais fácil de construir funções de densidade bidimensional é a partir de duas densidades unidimensionais f_1 e f_2 e definir a função f através de

$$(11) \qquad f(x,y) = f_1(x)f_2(y), \qquad -\infty < x, y < \infty.$$

Então, f será uma função de densidade bidimensional já que ela é claramente não-negativa e

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \int_{-\infty}^{\infty} f_1(x)\, dx \int_{-\infty}^{\infty} f_2(y)\, dy = 1.$$

Se as variáveis aleatórias X e Y têm f como sua densidade conjunta, então X e Y são independentes e têm densidades marginais $f_X = f_1$ e $f_Y = f_2$.

Como uma ilustração de (11), suponha que f_1 e f_2 são ambas densidade normal padrão $n(0,1)$. Então f é dada por

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

ou

$$f(x,y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}, \quad -\infty < x, y < \infty. \tag{12}$$

A densidade dada em (12) chama-se *densidade normal bidimensional padrão*. No exemplo seguinte modificaremos um pouco o segundo membro de (12) para obter uma função de densidade conjunta que corresponde ao caso em que as duas variáveis aleatórias tendo densidades marginais normais são dependentes.

Exemplo 2. Suponha que X e Y têm a função de densidade conjunta f dada por

$$f(x,y) = ce^{-(x^2 - xy + y^2)/2}, \quad -\infty < x, y < \infty,$$

onde c é uma constante positiva que será determinada no decorrer de nossa discussão. Inicialmente "completamos o quadrado" em termos envolvendo y e reescrevemos f como

$$f(x,y) = ce^{-[(y - x/2)^2 + 3x^2/4]/2}, \quad -\infty < x, y < \infty,$$

e então observamos que

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = ce^{-3x^2/8} \int_{-\infty}^{\infty} e^{-(y-x/2)^2/2}\, dy.$$

Fazendo a mudança de variável $u = y - x/2$, vemos que

$$\int_{-\infty}^{\infty} e^{-(y-x/2)^2/2}\, dy = \int_{-\infty}^{\infty} e^{-u^2/2}\, du = \sqrt{2\pi}.$$

Conseqüentemente

$$f_X(x) = c\sqrt{2\pi} e^{-3x^2/8}.$$

Agora torna-se claro que f_X é a densidade normal $n(0, \sigma^2)$ com $\sigma^2 = 4/3$ e portanto

$$c\sqrt{2\pi} = \frac{1}{\sigma\sqrt{2\pi}} = \frac{\sqrt{3}}{2\sqrt{2\pi}}$$

ou $c = \sqrt{3/4\pi}$. Conseqüentemente

$$f(x,y) = \frac{\sqrt{3}}{4\pi} e^{-(x^2 - xy + y^2)/2}, \quad -\infty < x, y < \infty. \tag{13}$$

Os cálculos acima mostram que f_X é a densidade normal $n(0, 4/3)$. Analogamente podemos mostrar que f_Y é também $n(0, 4/3)$. Como $f(x, y) \neq f_X(x)\, f_Y(y)$, é claro que X e Y são dependentes.

6.2. DISTRIBUIÇÃO DE SOMAS E QUOCIENTES

Sejam X e Y variáveis aleatórias tendo uma densidade conjunta f. Em muitos contextos temos uma variável aleatória Z definida em termos de X e Y e desejamos determinar a densidade de Z. Suponha que Z seja dada por $Z = \varphi(X, Y)$, onde φ é uma função real cujo domínio contém o contradomínio de X e Y. Para um z fixo o evento $\{Z \leqslant z\}$ é equivalente ao evento $\{(X, Y) \in A_z\}$, onde A_z é o subconjunto de R^2 definido por

$$A_z = \{(x,y) \mid \varphi(x,y) \leqslant z\}.$$

Assim

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P((X, Y) \in A_z) \\ &= \iint_{A_z} f(x, y)\, dx\, dy. \end{aligned}$$

Se pudermos obter uma função não-negativa g tal que

$$\iint_{A_z} f(x, y)\, dx\, dy = \int_{-\infty}^{z} g(v)\, dv, \qquad -\infty < z < \infty,$$

então g será necessariamente uma densidade de Z. Usaremos este método para obter as densidades de $X + Y$ e Y/X.

6.2.1. DISTRIBUIÇÃO DE SOMAS.

Seja $Z = X + Y$. Então

$$A_z = \{(x,y) \mid x + y \leqslant z\}$$

é simplesmente o semi-plano à esquerda inferior da reta $x + y = z$ como mostra a Figura 1. Assim

$$F_Z(z) = \iint_{A_z} f(x, y)\, dx\, dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-x} f(x, y)\, dy\right) dx.$$

Fazendo a mudança de variável $y = v - x$ na integral interna temos

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} f(x, v - x)\, dv\right) dx \\ &= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} f(x, v - x)\, dx\right) dv, \end{aligned}$$

onde invertemos a ordem de integração. Assim a densidade de $Z = X + Y$ é dada por

$$(14) \qquad f_{X+Y}(z) = \int_{-\infty}^{\infty} f(x, z - x)\, dx, \qquad -\infty < z < \infty.$$

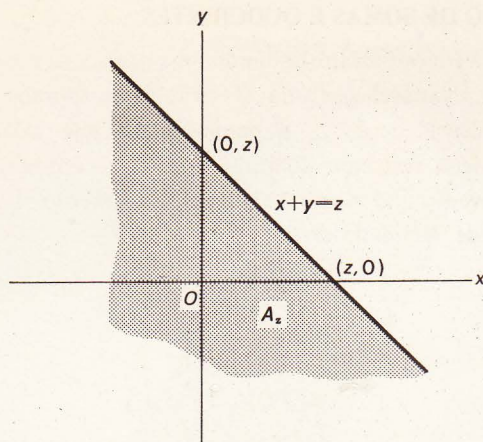



Figura 1

Na maioria das aplicações de (14), X e Y são independentes e pode-se reescrever (14) como

$$(15) \qquad f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx, \qquad -\infty < z < \infty.$$

Se X e Y forem variáveis aleatórias independentes não-negativas, então $f_{X+Y}(z) = 0$ para $z \leqslant 0$ e

$$(16) \qquad f_{X+Y}(z) = \int_0^z f_X(x) f_Y(z-x)\, dx, \qquad 0 < z < \infty.$$

O segundo membro de (15) sugere um método de obter densidades. Dadas duas densidades unidimensionais f e g, a função h definida por

$$h(z) = \int_{-\infty}^{\infty} f(x) g(z-x)\, dx, \qquad -\infty < z < \infty,$$

é uma função de densidade unidimensional que se chama *convolução* de f e g. Assim a densidade da soma de duas variáveis aleatórias independentes é a convolução das densidades individuais.

Exemplo 3. Sejam X e Y variáveis aleatórias independentes, cada uma com densidade exponencial de parâmetro λ. Obtenha a distribuição de $X + Y$.

A densidade de X é dada por $f_X(x) = \lambda e^{-\lambda x}$ para $x \geqslant 0$ e $f_X(x) = 0$ para $x < 0$. A densidade de Y é a mesma. Assim $f_{X+Y}(z) = 0$ para $z \leqslant 0$ e, de acordo com (16), para $z > 0$

$$\begin{aligned} f_{X+Y}(z) &= \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)}\, dx \\ &= \lambda^2 e^{-\lambda z} \int_0^z dx = \lambda^2 z e^{-\lambda z}. \end{aligned}$$

Vemos que $X + Y$ tem a densidade gama $\Gamma(2, \lambda)$.

Exemplo 4. Suponha que X e Y se distribuem independente e uniformemente em $(0, 1)$. Obtenha a densidade de $X + Y$.

A densidade de X é dada por $f_X(x) = 1$ para $0 < x < 1$ e $f_X(x) = 0$ para outros valores de x. A densidade de Y é a mesma. Assim $f_{X+Y}(z) = 0$ para $z \leqslant 0$. Para $z > 0$ aplicamos (16). O integrando $f_X(x) f_Y(z - x)$ assume apenas os valores de 0 e 1. Ele assume o valor 1 se x e z são tais que $0 \leqslant x \leqslant 1$ e $0 \leqslant z - x \leqslant 1$. Se $0 \leqslant z \leqslant 1$, o integrando tem valor 1 no intervalo $0 \leqslant x \leqslant z$ e zero caso contrário. Portanto obtemos de (16) que

$$f_{X+Y}(z) = z, \qquad 0 \leqslant z \leqslant 1.$$

Se $1 < z \leqslant 2$, o integrando tem valor 1 no intervalo $z - 1 \leqslant x \leqslant 1$ e zero caso contrário. Assim, de acordo com (16)

$$f_{X+Y}(z) = 2 - z, \qquad 1 < z \leqslant 2.$$

Se $2 < z < \infty$, o integrando de (16) é identicamente zero e portanto

$$f_{X+Y}(z) = 0, \qquad 2 < z < \infty.$$

Em resumo

$$f_{X+Y}(z) = \begin{cases} z, & 0 \leq z \leq 1, \\ 2 - z, & 1 < z \leq 2, \\ 0, & \text{para outros valores de } x. \end{cases}$$

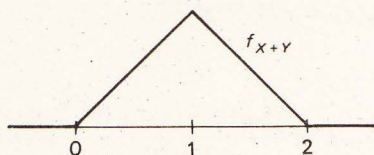



Figura 2

O gráfico de f é dado na Figura 2. Pode-se também obter a densidade de $X + Y$ determinando a área do conjunto

$$A_z = \{(x,y) \mid 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1 \quad \text{e} \quad x + y \leqslant z\}$$

(ver Figura 3) e diferenciando o resultado em relação a z.

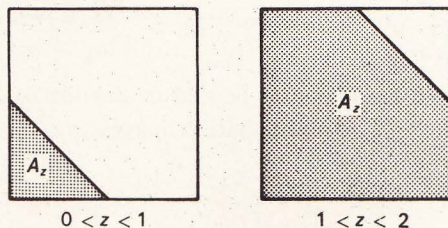



Figura 3

O Exemplo 3 tem uma importante generalização, que se pode formular como segue.

Teorema 1. Sejam X e Y variáveis aleatórias independentes tais que X tem a densidade gama $\Gamma(\alpha_1, \lambda)$ e Y tem a densidade gama $\Gamma(\alpha_2, \lambda)$. Então $X + Y$ têm a densidade gama

$$\Gamma(\alpha_1 + \alpha_2, \lambda).$$

Demonstração. Observamos que X e Y são variáveis aleatórias positivas e que

$$f_X(x) = \frac{\lambda^{\alpha_1} x^{\alpha_1 - 1} e^{-\lambda x}}{\Gamma(\alpha_1)}, \quad x > 0,$$

e

$$f_Y(y) = \frac{\lambda^{\alpha_2} y^{\alpha_2 - 1} e^{-\lambda y}}{\Gamma(\alpha_2)}, \quad y > 0.$$

Assim $f_{X+Y}(z) = 0$ para $z \leqslant 0$ e, de acordo com (16), para $z > 0$

$$f_{X+Y}(z) = \frac{\lambda^{\alpha_1 + \alpha_2} e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^z x^{\alpha_1 - 1}(z - x)^{\alpha_2 - 1}\, dx.$$

Fazendo a mudança de variável $x = zu$ (com $dx = z\, du$) na integral acima, obtemos

$$(17) \qquad f_{X+Y}(z) = c\lambda^{\alpha_1 + \alpha_2} z^{\alpha_1 + \alpha_2 - 1} e^{-\lambda z}, \quad z > 0,$$

onde

$$(18) \qquad c = \frac{\int_0^1 u^{\alpha_1 - 1}(1 - u)^{\alpha_2 - 1}\, du}{\Gamma(\alpha_1)\Gamma(\alpha_2)}.$$

Pode-se determinar a constante c pelo fato de que a integral de f_{X+Y} é 1. De (17) e da definição de densidades gama torna-se claro que f_{X+Y} deve ser a densidade gama $\Gamma(\alpha_1 + \alpha_2, \lambda)$ como foi afirmado.

De (17) e da definição de densidade gama vemos também que $c = 1/\Gamma(\alpha_1 + \alpha_2)$. Isto juntamente com (18) permite-nos determinar a integral definida que aparece em (18) em termos de função gama:

$$(19) \qquad \int_0^1 u^{\alpha_1 - 1}(1 - u)^{\alpha_2 - 1}\, du = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}.$$

Esta fórmula permite-nos definir outra família de densidades com dois parâmetros chamadas densidades *Beta*. A densidade Beta de parâmetros α_1 e α_2 é definida por

$$(20) \qquad f(x) = \begin{cases} \dfrac{\Gamma(\alpha_1 + \alpha_2) x^{\alpha_1 - 1}(1 - x)^{\alpha_2 - 1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}, & 0 < x < 1, \\ 0, & \text{para outros valores de } x. \end{cases}$$

A razão desta terminologia é que a função de α_1 e α_2 definida por

$$B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}, \qquad 0 < \alpha_1, \alpha_2 < \infty,$$

chama-se função beta.

Nossa aplicação final da fórmula de convolução será à variáveis aleatórias normalmente distribuídas.

Teorema 2. Sejam X e Y variáveis aleatórias independentes com densidades normais $n(\mu_1, \sigma_1^2)$ e $n(\mu_2, \sigma_2^2)$, respectivamente. Então $X + Y$ tem a densidade normal

$$n(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Demonstração. Suponhamos inicialmente que $\mu_1 = \mu_2 = 0$. Então

$$f_X(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-x^2/2\sigma_1^2}, \qquad -\infty < x < \infty,$$

e

$$f_Y(y) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-y^2/2\sigma_2^2}, \qquad -\infty < y < \infty.$$

Assim de acordo com (15)

$$f_{X+Y}(z) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_1^2} + \frac{(z - x)^2}{\sigma_2^2}\right)\right] dx.$$

Infelizmente a determinação desta integral requer alguns cálculos laboriosos (que não são suficientemente importantes para justificar seu domínio). Uma das maneiras de proceder é fazer inicialmente a mudança de variável

$$u = \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1 \sigma_2} x.$$

Após algumas transformações algébricas obtemos

$$f_{X+Y}(z) = \frac{1}{2\pi\sqrt{\sigma_1^2 + \sigma_2^2}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(u^2 - \frac{2uz\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}} + \frac{z^2}{\sigma_2^2}\right)\right] du.$$

A seguir completamos o quadro em u e observamos que

$$u^2 - \frac{2uz\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}} + \frac{z^2}{\sigma_2^2} = \left(u - \frac{z\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}}\right)^2 + \frac{z^2}{\sigma_1^2 + \sigma_2^2}.$$

Então fazendo uma segunda mudança de variável

$$v = u - \frac{z\sigma_1}{\sigma_2\sqrt{\sigma_1^2 + \sigma_2^2}},$$

vemos que

$$f_{X+Y}(z) = \frac{e^{-z^2/2(\sigma_1^2+\sigma_2^2)}}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} \int_{-\infty}^{\infty} \frac{e^{-v^2/2}}{\sqrt{2\pi}} dv$$

$$= \frac{e^{-z^2/2(\sigma_1^2+\sigma_2^2)}}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}},$$

que é simplesmente a densidade normal $n(0, \sigma_1^2 + \sigma_2^2)$.

No caso geral $X - \mu_1$ e $Y - \mu_2$ são independentes e têm densidades normais $n(0, \sigma_1^2)$ e $n(0, \sigma_2^2)$, respectivamente. Assim de acordo com o caso especial acima, $(X - \mu_1) + (Y - \mu_2) = X + Y - (\mu_1 + \mu_2)$ têm a densidade normal $n(0, \sigma_1^2 + \sigma_2^2)$, e portanto $X + Y$ têm a densidade normal

$$n(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

como afirmamos.

A demonstração acima é elementar porém trabalhosa. Uma demonstração menos computacional envolvendo técnicas mais avançadas será dada na Seção 8.3. Outra demonstração é indicada no Exercício 36 no final deste capítulo.

Exemplo 5. Sejam X e Y variáveis aleatórias independentes cada uma com densidade normal $n(0, \sigma^2)$. Obtenha as densidades de $X + Y$ e $X^2 + Y^2$.

Pelo Teorema 2 vemos de imediato que $X + Y$ têm a densidade normal $n(0, 2\sigma^2)$. De acordo com o Exemplo 12 do Capítulo 5, X^2 e Y^2 têm individualmente a densidade gama $\Gamma(1/2, 1/2\sigma^2)$. Vê-se facilmente que X^2 e Y^2 são independentes. Assim, pelo Teorema 1, $X^2 + Y^2$ têm a densidade gama $\Gamma(1, 1/2\sigma^2)$ que é também a densidade exponencial de parâmetro $1/2\sigma^2$.

6.2.2. DISTRIBUIÇÃO DE QUOCIENTES.* Suponha que X e Y representam variáveis aleatórias com densidade conjunta f. Derivaremos a seguir uma fórmula para a densidade da variável aleatória $Z = Y/X$. A figura 4 mostra o conjunto

$$A_z = \{(x,y) \mid y/x \leq z\}$$

Se $x < 0$, então $y/x \leq z$ se, e somente se, $y \geq xz$. Assim

$$A_z = \{(x,y) \mid x < 0 \text{ e } y \geq xz\} \cup \{(x,y) \mid x > 0 \text{ e } y \leq xz\}.$$

Conseqüentemente

$$F_{Y/X}(z) = \iint_{A_z} f(x, y)\, dx\, dy$$

$$= \int_{-\infty}^{0} \left(\int_{xz}^{\infty} f(x, y)\, dy \right) dx + \int_{0}^{\infty} \left(\int_{-\infty}^{xz} f(x, y)\, dy \right) dx.$$

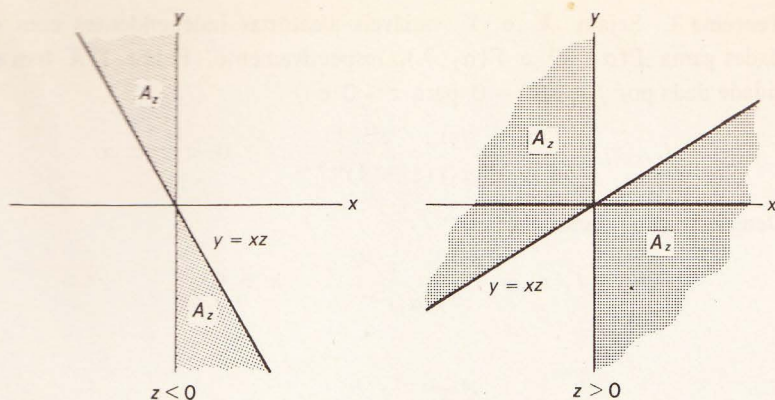



Figura 4

Fazendo a mudança de variável $y = xv$ (com $dy = x\, dv$) nas integrais internas obtemos

$$
\begin{aligned}
F_{Y/X}(z) &= \int_{-\infty}^{0} \left(\int_{z}^{-\infty} x f(x, xv)\, dv \right) dx \\
&\quad + \int_{0}^{\infty} \left(\int_{-\infty}^{z} x f(x, xv)\, dv \right) dx \\
&= \int_{-\infty}^{0} \left(\int_{-\infty}^{z} (-x) f(x, xv)\, dv \right) dx \\
&\quad + \int_{0}^{\infty} \left(\int_{-\infty}^{z} x f(x, xv)\, dv \right) dx \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z} |x| f(x, xv)\, dv \right) dx.
\end{aligned}
$$

Invertendo a ordem de integração vemos que

(21) $$F_{Y/X}(z) = \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} |x| f(x, xv)\, dx \right) dv, \qquad -\infty < z < \infty.$$

Segue-se de (21) que Y/X têm a densidade $f_{Y/X}$ dada por

(22) $$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f(x, xz)\, dx, \qquad -\infty < z < \infty.$$

No caso especial em que X e Y são variáveis aleatórias independentes positivas, (22) reduz-se a $f_{Y/X}(z) = 0$ para $z \leqslant 0$ e

(23) $$f_{Y/X}(z) = \int_{0}^{\infty} x f_X(x) f_Y(xz)\, dx, \qquad 0 < z < \infty.$$

O teorema a seguir é uma aplicação direta de (23).

Teorema 3. Sejam X e Y variáveis aleatórias independentes com densidades gama $\Gamma(\alpha_1, \lambda)$ e $\Gamma(\alpha_2, \lambda)$, respectivamente. Então Y/X tem a densidade dada por $f_{Y/X}(z) = 0$ para $z \leqslant 0$ e

$$f_{Y/X}(z) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{z^{\alpha_2 - 1}}{(z + 1)^{\alpha_1 + \alpha_2}}, \qquad 0 < z < \infty. \tag{24}$$

Demonstração. Lembre-se que

$$f_X(x) = \frac{\lambda^{\alpha_1} x^{\alpha_1 - 1} e^{-\lambda x}}{\Gamma(\alpha_1)}, \qquad x > 0,$$

e

$$f_Y(y) = \frac{\lambda^{\alpha_2} y^{\alpha_2 - 1} e^{-\lambda y}}{\Gamma(\alpha_2)}, \qquad y > 0.$$

Pode-se aplicar (23), de modo que para $0 < z < \infty$,

$$\begin{aligned} f_{Y/X}(z) &= \frac{\lambda^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty x x^{\alpha_1 - 1} e^{-\lambda x} (xz)^{\alpha_2 - 1} e^{-\lambda xz}\, dx \\ &= \frac{\lambda^{\alpha_1 + \alpha_2} z^{\alpha_2 - 1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^\infty x^{\alpha_1 + \alpha_2 - 1} e^{-x\lambda(z+1)}\, dx. \end{aligned}$$

De acordo com a Equação (34) do Capítulo 5

$$\int_0^\infty x^{\alpha_1 + \alpha_2 - 1} e^{-x\lambda(z+1)}\, dx = \frac{\Gamma(\alpha_1 + \alpha_2)}{(\lambda(z + 1))^{\alpha_1 + \alpha_2}}.$$

Conseqüentemente a fórmula (24) é válida como foi afirmado.

Uma vez que (24) define uma função de densidade, vemos que para $\alpha_1, \alpha_2 > 0$

$$\int_0^\infty z^{\alpha_2 - 1}(z + 1)^{-(\alpha_1 + \alpha_2)}\, dz = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}.$$

Exemplo 6. Sejam X e Y variáveis aleatórias independentes, cada uma com densidade normal $n(0, \sigma^2)$. Obtenha a densidade de Y^2/X^2.

As variáveis aleatórias são as mesmas do Exemplo 5. Assim X^2 e Y^2 são novamente independentes e têm a densidade gama $\Gamma(1/2, 1/2\sigma^2)$. O Teorema 3 é então aplicável e Y^2/X^2 têm a densidade f_{Y^2/X^2} dada por $f_{Y^2/X^2}(z) = 0$ para $z \leqslant 0$ e

$$\begin{aligned} f_{Y^2/X^2}(z) &= \frac{\Gamma(1)}{\Gamma(1/2)\Gamma(1/2)} \frac{z^{-1/2}}{(z + 1)} \\ &= \frac{1}{\pi(z + 1)\sqrt{z}}, \qquad 0 < z < \infty. \end{aligned}$$

(Lembramos da Equação (35) do Capítulo 5 que $\Gamma(1/2) = \sqrt{\pi}$). A título de exercício deixamos para o leitor mostrar que, sob as mesmas condições, tanto Y/X como $Y/|X|$ têm a densidade de Cauchy.

6.3. DENSIDADES CONDICIONAIS

Para motivar a definição de densidades condicionais de variáveis aleatórias contínuas, discutiremos inicialmente as variáveis aleatórias discretas. Sejam X e Y duas variáveis aleatórias discretas tendo densidade conjunta f. Se x é um valor possível de X, então

$$P(Y=y \mid X=x) = \frac{P(X=x, Y=y)}{P(X=x)} = \frac{f(x,y)}{f_X(x)} .$$

A função $f_{Y/X}$ definida por

$$(25) \qquad f_{Y|X}(y \mid x) = \begin{cases} \dfrac{f(x, y)}{f_X(x)}, & f_X(x) \neq 0, \\ 0, & f_X(x) = 0, \end{cases}$$

chama-se densidade condicional de Y dado X. Para qualquer valor possível x de X,

$$\sum_y f_{Y|X}(y \mid x) = \frac{\sum_y f(x, y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1,$$

de modo que para qualquer x nestas condições $f_{Y/X}(y/x)$ define uma função de densidade discreta de y, conhecida como densidade condicional de Y dado $X=x$. No caso discreto as densidades condicionais não envolvem nenhum conceito novo.

Entretanto se X é uma variável aleatória contínua, então $P(X = x) = 0$ para todo x, de modo que $P(Y = y/X = x)$ é sempre indefinido. Neste caso, qualquer definição de densidades condicionais envolve necessariamente um conceito novo. A maneira mais simples de definir densidades condicionais de variáveis aleatórias contínuas é uma analogia com a Fórmula (25) do caso discreto.

Definição 1. Sejam X e Y variáveis aleatórias contínuas tendo densidade conjunta f. Define-se a densidade condicional $f_{Y/X}$ através de

$$(26) \qquad f_{Y|X}(y \mid x) = \begin{cases} \dfrac{f(x, y)}{f_X(x)}, & 0 < f_X(x) < \infty, \\ 0, & \text{para outros valores de } x \text{ e } y. \end{cases}$$

Segue-se imediatamente desta definição que, como uma função de y, $f_{Y/X}(y/x)$ é uma densidade sempre que $0 < f_X(x) < \infty$ (novamente denominada densidade condicional de Y dado $X = x$). Pode-se utilizar as densidades condicionais para definir probabilidades condicionais. Assim definimos

$$(27) \qquad P(a \leq Y \leq b \mid X = x) = \int_a^b f_{Y|X}(y \mid x)\, dy, \qquad a \leq b.$$

Alternativamente poderíamos tentar definir a probabilidade condicional de (27) através do seguinte limite:

$$(28) \qquad P(a \leq Y \leq b \mid X = x)$$

$$= \lim_{h \downarrow 0} P(a \leq Y \leq b \mid x - h \leq X \leq x + h).$$

Pode-se reescrever o segundo membro de (28) em termos de f como

$$\lim_{h \downarrow 0} \frac{\int_{x-h}^{x+h} \left(\int_a^b f(u, y)\, dy \right) du}{\int_{x-h}^{x+h} \left(\int_{-\infty}^{\infty} f(u, y)\, dy \right) du} = \lim_{h \downarrow 0} \frac{(1/2h) \int_{x-h}^{x+h} \left(\int_a^b f(u, y)\, dy \right) du}{(1/2h) \int_{x-h}^{x+h} f_X(u)\, du}.$$

Se

$$\int_a^b f(u, y)\, dy$$

é contínua em u para $u = x$, o numerador do último limite converge para

$$\int_a^b f(x, y)\, dy$$

quando $h \downarrow 0$. Se f_X é contínua no ponto x, o denominador converge para $f_X(x)$ quando $h \downarrow 0$. Sob a condição adicional de que $f_X(x) \neq 0$, somos levados de (28) a

$$P(a \leq Y \leq b \mid X = x) = \frac{\int_a^b f(x, y)\, dy}{f_X(x)},$$

que concorda com (27). Em resumo, definimos densidades condicionais e probabilidades condicionais no caso contínuo em analogia com o caso discreto. Observamos também que, sob restrições adicionais, um processo envolvendo limites produziria a mesma definição de probabilidades condicionais. Acontece que é bastante difícil trabalhar com tais processos, de modo que os mesmos não mais serão usados.

Segue-se imediatamente da definição de funções de densidades condicionais que

$$f(x,y) = f_X(x) f_{Y \mid X}(y \mid x), \quad -\infty < x, y < \infty. \tag{29}$$

Se X e Y são independentes e

$$f(x,y) = f_X(x) f_Y(y), \quad -\infty < x, y < \infty, \tag{30}$$

então

$$f_{Y \mid X}(y \mid x) = f_Y(y), \quad 0 < f_X(x) < \infty \quad \text{e} \quad -\infty < y < \infty. \tag{31}$$

Reciprocamente, se (31) é verdadeiro, segue-se de (29) que (30) é verdadeiro e X e Y são independentes. Assim (31) é uma condição necessária e suficiente para que duas variáveis aleatórias tendo uma densidade conjunta sejam independentes.

Exemplo 7. Suponha que X e Y têm a densidade bidimensional f dada pela Fórmula (13), ou seja

$$f(x, y) = \frac{\sqrt{3}}{4\pi} e^{-(x^2 - xy + y^2)/2}, \quad -\infty < x, y < \infty.$$

Então, como vimos no Exemplo 2, X tem densidade normal $n(0, 4/3)$. Assim para $-\infty < x, y < \infty$

$$f_{Y|X}(y \mid x) = \frac{\dfrac{\sqrt{3}}{4\pi} e^{-(x^2-xy+y^2)/2}}{\dfrac{\sqrt{3}}{2\sqrt{2\pi}} e^{-3x^2/8}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-(y-x/2)^2/2}.$$

Em outras palavras, a densidade condicional de Y dado $X = x$ é a densidade normal $n(x/2, 1)$.

Até aqui partimos de densidades conjuntas e através delas construímos densidades marginais e densidades condicionais. Em certas situações podemos inverter este procedimento e começar com densidades marginais e densidades condicionais, usando-as para construir densidades conjuntas.

Exemplo 8. Seja X uma variável aleatória uniformemente distribuída em (0, 1) e seja Y uma variável aleatória uniformemente distribuída em $(0, X)$. Obtenha a densidade conjunta de X e Y e a densidade marginal de Y.

Pelo enunciado do problema vemos que a densidade marginal de X é dada por

$$f_X(x) = \begin{cases} 1 & \text{para} \quad 0<x<1. \\ 0 & \text{para outros valores de } x \end{cases}$$

A densidade de Y dado $X=x$ é uniforme em $(0,x)$, de modo que

$$f_{Y|X}(y \mid x) = \begin{cases} 1/x & \text{para} \quad 0<y<x<1, \\ 0, & \text{para outros valores de } x \text{ e } y \end{cases}$$

A densidade conjunta de X e Y é dada por

$$f(x,y) = \begin{cases} 1/x & \text{para} \quad 0<y<x<1, \\ 0 & \text{para outros valores de } x \text{ e } y \end{cases}$$

A densidade marginal de Y é

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_y^1 \frac{1}{x}\, dx = -\log y, \qquad 0 < y < 1,$$

e $f_Y(y) = 0$ para outros valores de y.

6.3.1. REGRA DE BAYES. Naturalmente podemos inverter os papéis de X e Y e definir a densidade condicional de X dado $Y=y$ através da fórmula

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)}, \quad 0<f_Y(y)<\infty. \tag{32}$$

Uma vez que

$$f(x, y) = f_X(x) f_{Y|X}(y \mid x)$$

e

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y \mid x)\, dx,$$

podemos reescrever (32) como

$$(33) \qquad f_{X|Y}(x \mid y) = \frac{f_X(x) f_{Y|X}(y \mid x)}{\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y \mid x)\, dx}.$$

Esta fórmula é o análogo contínuo da famosa regra de Bayes discutida no Capítulo 1.

Nos Capítulos 3 e 4 consideramos variáveis aleatórias X e Y que eram ambas discretas. Até aqui no Capítulo 6 consideramos principalmente variáveis aleatórias X e Y que são ambas contínuas. Existem casos em que estamos interessados simultaneamente em variáveis aleatórias discretas e contínuas. Deve ficar claro para o leitor como podemos modificar nossa discussão para incluir esta possibilidade. Algumas das aplicações mais interessantes de (33) são deste tipo.

Exemplo 9. Suponha que o número de acidentes automobilísticos em que se envolve um motorista durante um ano é uma variável aleatória Y tendo uma distribuição de Poisson de parâmetro λ, onde λ depende do motorista. Se escolhemos um motorista ao acaso de uma certa população, podemos supor que λ varia e define uma variável aleatória contínua Λ tendo uma densidade f_Λ. A densidade condicional de Y dado $\Lambda = \lambda$ é uma densidade de Poisson de parâmetro λ dada por

$$f_{Y|\Lambda}(y \mid \lambda) = \begin{cases} \dfrac{\lambda^y e^{-\lambda}}{y!} & \text{para} \quad y = 0, 1, 2, \ldots, \\ 0, & \text{para outros valores de } y \end{cases}$$

Assim a densidade conjunta de Λ e Y é

$$f(\lambda, y) = \begin{cases} \dfrac{f_\Lambda(\lambda) \lambda^y e^{-\lambda}}{y!} & \text{para} \quad y = 0, 1, 2, \ldots, \\ 0, & \text{para outros valores de } y \end{cases}$$

Em geral não podemos obter uma fórmula simples para a densidade marginal de Y ou a densidade condicional de Λ dado $Y = y$ já que não podemos determinar as integrais necessárias. Entretanto podemos obter fórmulas simples no caso especial em que f é uma densidade gama $\Gamma(\alpha, \beta)$. de modo que

$$f_\Lambda(\lambda) = \begin{cases} \dfrac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)} & \text{para } \lambda > 0 \\ 0, & \text{para outros valores de } y. \end{cases}$$

Neste caso,

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f(\lambda, y)\, d\lambda \\
&= \int_0^{\infty} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)} \frac{\lambda^y e^{-\lambda}}{y!}\, d\lambda \\
&= \frac{\beta^\alpha}{y!\Gamma(\alpha)} \int_0^{\infty} \lambda^{\alpha+y-1} e^{-\lambda(\beta+1)}\, d\lambda \\
&= \frac{\Gamma(\alpha + y)\beta^\alpha}{y!\Gamma(\alpha)(\beta + 1)^{\alpha+y}}.
\end{aligned}$$

O valor da última integral foi obtido usando a fórmula (34) do Capítulo 5. Deixamos para o leitor a título de exercício mostrar que f_Y é a densidade binomial negativa de parâmetros α e $p = \beta/(1 + \beta)$. Temos também que para $\lambda > 0$ e y inteiro não-negativo,

$$\begin{aligned}
f_{\Lambda|Y}(\lambda \mid y) &= \frac{f(\lambda, y)}{f_Y(y)} \\
&= \frac{\beta^\alpha \lambda^{\alpha+y-1} e^{-\lambda(\beta+1)} y!\Gamma(\alpha)(\beta + 1)^{\alpha+y}}{\Gamma(\alpha) y!\Gamma(\alpha + y)\beta^\alpha} \\
&= \frac{(\beta + 1)^{\alpha+y} \lambda^{\alpha+y-1} e^{-\lambda(\beta+1)}}{\Gamma(\alpha + y)},
\end{aligned}$$

que nos diz que a densidade condicional de Λ dado $Y = y$ é a densidade gama $\Gamma(\alpha + y, \beta + 1)$. Se alguém de uma companhia de seguros desejasse resolver um problema deste tipo, possivelmente tentaria aproximar a densidade real f_Λ para uma densidade $\Gamma(\alpha, \beta)$, onde α e β seriam determinados de modo que a aproximação fosse a melhor possível.

6.4. PROPRIEDADES DE DISTRIBUIÇÕES MULTIDIMENSIONAIS

Pode-se estender imediatamente para n variáveis aleatórias os conceitos até aqui discutidos em relação a duas variáveis aleatórias X e Y. Indicaremos brevemente nesta seção como isto pode ser feito.

Sejam n variáveis aleatórias $X_1, \ldots, X_n$ definidas em um espaço comum de probabilidade. Define-se sua *função de distribuição conjunta* F através de

$$F(x_1, \ldots, x_n) = P(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n), \qquad -\infty < x_1, \ldots, x_n < \infty.$$

Definem-se as funções de distribuição marginal F_{X_m}, $m = 1, \ldots, n$ através de

$$F_{X_m}(x_m) = P(X_m \leqslant x_m), \qquad -\infty < x_m < \infty.$$

Pode-se obter o valor de $F_{X_m}(x_m)$ a partir de F fazendo $x_1, \ldots, x_{m-1}, x_{m+1}, \ldots, x_n$ tenderem todos a $+\infty$.

Diz-se que uma função não-negativa f é uma função de densidade conjunta (em relação à integração) para a função de distribuição conjunta F, ou para as variáveis aleatórias $X_1, \ldots, X_n$ se

$$(34) \qquad F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(u_1, \ldots, u_n)\, du_1 \cdots du_n,$$
$$-\infty < x_1, \ldots, x_n < \infty.$$

Sob algumas condições brandas adicionais, a equação

$$f(x_1, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \ldots, x_n)$$

é válida nos pontos de continuidade de F. Se (34) for válido e A for um subconjunto qualquer de R^n do tipo considerado em cálculo, então

$$P((X_1, \ldots, X_n) \in A) = \int \cdots \int_A f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

Em particular

$$(35) \qquad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$$

e se $a_m \leqslant b_m$ para $m = 1, \ldots, n$, então

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n)$$
$$= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

A variável aleatória X_m tem densidade marginal f_{X_m} obtida pela integração de f sobre as $n-1$ variáveis restantes. Por exemplo

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_1\, dx_3 \cdots dx_n.$$

Em geral diz-se que as variáveis aleatórias $X_1, \ldots, X_n$ são independentes sempre que $a_m \leqslant b_m$, para $m = 1, , \ldots, n$, então

$$P(a_1 < X_1 \leqslant b_1, \ldots, a_n < X_n \leqslant b_n)$$
$$= P(a_1 < X_1 \leqslant b_1) \cdots P(a_n < X_n \leqslant b_n).$$

Uma condição necessária e suficiente de independência é

$$F(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n), \qquad -\infty < x_1, \ldots, x_n < \infty.$$

A necessidade é óbvia, mas a parte da suficiência para $n > 2$ depende de artifícios peculiares e não será demonstrada aqui. Se F tem densidade f, então $X_1, \ldots, X_n$ são independentes se, e somente se, pudermos escolher f de modo que

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n), \quad -\infty < x_1, \ldots, x_n < \infty.$$

Pode-se também definir diretamente uma densidade n-dimensional como uma função não-negativa em R^n para a qual (35) é válido. A maneira mais simples de construir densidades n-dimensionais é a partir de n densidades unidimensionais $f_1, \ldots, f_n$ e definir f através de

$$(36) \qquad f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n), \qquad -\infty < x_1, \ldots, x_n < \infty.$$

Se $X_1, \ldots, X_n$ são variáveis aleatórias cuja densidade conjunta f é dada por (36), então $X_1, \ldots, X_n$ são independentes e X_m tem densidade marginal f_m.

Exemplo 10. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes, cada uma com densidade exponencial de parâmetro λ. Obtenha a densidade conjunta de $X_1, \ldots, X_n$.

A densidade de X_m é dada por

$$f_{X_m}(x_m) = \begin{cases} \lambda e^{-\lambda x_m} & \text{para} \quad 0 < x_m < \infty, \\ 0, & \text{para outros valores de } x. \end{cases}$$

Assim f é dada por

$$f(x_1, \ldots, x_n) = \begin{cases} \lambda^n e^{-\lambda(x_1 + \cdots + x_n)} & \text{para} \quad x_1, \ldots, x_n > 0, \\ 0, & \text{para outros valores de } x_1, \ldots, x_n. \end{cases}$$

Para determinar a densidade de somas de n variáveis aleatórias independentes, bem como para muitos outros propósitos, necessitamos do seguinte resultado.

Teorema 4. Sejam n variáveis aleatórias independentes $X_1, \ldots, X_n$. Seja Y uma variável aleatória definida em termos de $X_1, \ldots, X_m$ e seja Z uma variável aleatória definida em termos de $X_{m+1}, \ldots, X_n$ (com $1 \leqslant m < n$) Então Y e Z são independentes.

A demonstração deste teorema não será apresentada já que ela envolve argumentos da teoria da medida.

Usando este teorema e um argumento envolvendo indução matemática, podemos estender os Teoremas 1 e 2 para somas de n variáveis aleatórias independentes como segue.

Teorema 5. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes tais que X_m tem densidade gama $\Gamma(\alpha_m, \lambda)$ para $m = 1, \ldots, n$. Então $X_1 + \cdots + X_n$ tem densidade gama $\Gamma(\alpha, \lambda)$, onde $\alpha = \alpha_1 + \cdots + \alpha_n$.

Lembre-se que a densidade exponencial de parâmetro λ é a densidade gama $\Gamma(1, \lambda)$. Assim, como um caso especial deste teorema, temos o seguinte corolário:

Se $X_1, \ldots, X_n$ são variáveis aleatórias independentes, cada uma com densidade exponencial de parâmetro λ, então $X_1 + \cdots + X_n$ têm densidade gama $\Gamma(n, \lambda)$.

Teorema 6. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes tais que X_m tem densidade normal $n(\mu_m, \sigma_m^2)$, $m = 1, \ldots, n$. Então $X_1 + \cdots + X_n$ têm densidade normal $n(\sigma, \sigma^2)$, onde

$$\mu = \mu_1 + \cdots + \mu_n \text{ e } \sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2.$$

Se $X_1, \ldots, X_n$ têm uma densidade conjunta f, então qualquer subcoleção destas variáveis aleatórias tem uma densidade conjunta que pode ser obtida integrando f sobre as variáveis restantes. Por exemplo, se $1 \leqslant m < n$,

$$f_{X_1,\ldots,X_m}(x_1, \ldots, x_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_{m+1} \cdots dx_n.$$

Pode-se também definir de maneira óbvia a densidade condicional de uma subcoleção de $X_1, \ldots, X_n$, dadas as variáveis restantes. Assim, a densidade condicional de $X_{m+1}, \ldots, X_n$ dados $X_1, \ldots, X_m$ é definida por

$$f_{X_{m+1},\ldots,X_n|X_1,\ldots,X_m}(x_{m+1}, \ldots, x_n \mid x_1, \ldots, x_m) = \frac{f(x_1, \ldots, x_n)}{f_{X_1,\ldots,X_m}(x_1, \ldots, x_m)},$$

onde f é a densidade conjunta de $X_1, \ldots, X_n$.

Freqüentemente expressa-se as densidades condicionais através de uma notação um tanto diferente. Por exemplo, sejam $n + 1$ variáveis aleatórias com densidade conjunta f. Então a densidade condicional de Y dados $X_1, \ldots, X_n$ é definida por

$$f_{Y|X_1,\ldots,X_n}(y \mid x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n, y)}{f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)}.$$

6.5 ESTATÍSTICOS DE ORDEM*

Sejam $U_1, \ldots, U_n$ variáveis aleatórias contínuas independentes, cada uma com função de distribuição F e função de densidade f. Sejam $X_1, \ldots, X_n$ as variáveis aleatórias obtidas representando por $X_1(\omega), \ldots, X_n(\omega)$ o conjunto $U_1(\omega), \ldots, U_n(\omega)$ ordenado em ordem crescente. Em particular, X_1 e X_n são definidas como sendo as funções

$$X_1(\omega) = \min(U_1(\omega), \ldots, U_n(\omega))$$

e

$$X_n(\omega) = \max(U_1(\omega), \ldots, U_n(\omega)).$$

A variável aleatória X_k chama-se k-ésimo *estatístico de ordem.* Outra variável correlata de interesse é a amplitude R definida como

$$R(\omega) = X_n(\omega) - X_1(\omega)$$
$$= \max(U_1(\omega), \ldots, U_n(\omega)) - \min(U_1(\omega), \ldots, U_n(\omega)).$$

Segue-se das hipóteses sobre $U_i, \ldots, U_n$ que os U_i são distintos com probabilidade 1 e portanto $X_1 < X_2 < \cdots < X_n$.

Para ilustrar numericamente estas definições, suponha que $U_1(\omega) = 4{,}8$; $U_2(\omega) = 3{,}5$ e $U_3(\omega) = 4{,}3$. Então $X_1(\omega) = 3{,}5$; $X_2(\omega) = 4{,}3$; $X_3(\omega) = 4{,}8$ e $R(\omega) = 1{,}3$.

Exemplo 11. Considere uma máquina tendo n componentes cujos tempos de falha satisfazem as hipóteses desta seção. Então X_k é o tempo que decorre até a falha de k componentes. Se a máquina como um todo falha tão logo que falhe um único componente, então $X_1 = \min(U_1, \ldots, U_n)$ é o tempo de falha da máquina. Se a máquina não falha até que todos os componentes falhem, então $X_n = \max(U_1, \ldots, U_n)$ é o tempo de falha da máquina.

Exemplo 12. Suponha que se produz n itens supostamente idênticos em uma corrida de uma linha de montagem e suponha que $U_1, \ldots, U_n$ representam os comprimentos dos n itens. Um inspetor poderia estar interessado no comprimento mínimo X_1 e no comprimento máximo X_n. Se os itens devem ser intercambiáveis, a variabilidade dos comprimentos deve ser mantida baixa. Uma possível medida desta variabilidade é a amplitude R dos comprimentos.

Passamos agora à obtenção da função de distribuição do k-ésimo estatístico de ordem X_k. Seja $-\infty < x < \infty$. A probabilidade de que exatamente j dos U_i estejam em $(-\infty, x]$ e $(n-j)$ estejam em (x, ∞) é

$$\binom{n}{j} F^j(x)(1 - F(x))^{n-j},$$

uma vez que é aplicável a distribuição binomial de parâmetros n e $p = F(x)$. O evento $\{X_k \leqslant x\}$ ocorre se, e somente se, k ou mais dos U_i estiverem em $(-\infty, x]$. Assim

(37) $$F_{X_k}(x) = P(X_k \leq x)$$
$$= \sum_{j=k}^{n} \binom{n}{j} F^j(x)(1 - F(x))^{n-j}, \qquad -\infty < x < \infty.$$

Em particular, pode-se escrever com muita simplicidade as funções de distribuição de X_n e X_1 como

$$F_{X_n}(x) = (F(x))^n, \qquad -\infty < x < \infty,$$

e

$$F_{X_1}(x) = 1 - (1 - F(x))^n, \qquad -\infty < x < \infty.$$

Para obter as funções de densidade correspondentes, devemos diferenciar essas expressões. Obtemos facilmente que

$$f_{X_n}(x) = nf(x)F^{n-1}(x), \qquad -\infty < x < \infty,$$

e

$$f_{X_1}(x) = nf(x)(1 - F(x))^{n-1}, \qquad -\infty < x < \infty.$$

A derivação correspondente para X_k em geral é um pouco mais trabalhosa. De (37),

$$\begin{aligned} f_{X_k}(x) &= \sum_{j=k}^{n} \frac{n!}{(j-1)!\,(n-j)!} f(x)F^{j-1}(x)(1 - F(x))^{n-j} \\ &\quad - \sum_{j=k}^{n-1} \frac{n!}{j!\,(n-j-1)!} f(x)F^{j}(x)(1 - F(x))^{n-j-1} \\ &= \sum_{j=k}^{n} \frac{n!}{(j-1)!\,(n-j)!} f(x)F^{j-1}(x)(1 - F(x))^{n-j} \\ &\quad - \sum_{j=k+1}^{n} \frac{n!}{(j-1)!\,(n-j)!} f(x)F^{j-1}(x)(1 - F(x))^{n-j} \end{aligned}$$

e por cancelação

$$(38) \qquad f_{X_k}(x) = \frac{n!}{(k-1)!\,(n-k)!} f(x)F^{k-1}(x)(1 - F(x))^{n-k}, \qquad -\infty < x < \infty.$$

Para obter a densidade da amplitude R, derivamos inicialmente a densidade conjunta de X_1 e X_n. Supomos que $n \geqslant 2$ (já que $R = 0$ se $n = 1$). Seja $x \leqslant y$. Então

$$\begin{aligned} P(X_1 > x, X_n \leqslant y) &= P(x < U_1 \leqslant y, \ldots, x < U_n \leqslant y) \\ &= (F(y) - F(x))^n, \end{aligned}$$

e naturalmente

$$P(X_n \leqslant y) = F^n(y).$$

Conseqüentemente

$$\begin{aligned} F_{X_1,X_n}(x, y) &= P(X_1 \le x, X_n \le y) \\ &= P(X_n \le y) - P(X_1 > x, X_n \le y) \\ &= F^n(y) - (F(y) - F(x))^n. \end{aligned}$$

A densidade conjunta é dada por

$$\begin{aligned} f_{X_1,X_n}(x, y) &= \frac{\partial^2}{\partial x\, \partial y} F_{X_1,X_n}(x, y) \\ &= n(n-1)f(x)f(y)(F(y) - F(x))^{n-2}, \qquad x \le y. \end{aligned}$$

É óbvio e demonstra-se facilmente que

$$f_{X_1, X_n}(x, y) = 0, \qquad x > y.$$

Modificando um pouco o argumento usado na Seção 6.2.1 para obter a densidade de uma soma, obtemos que a densidade de $R = X_n - X_1$ é dada por

$$f_R(r) = \int_{-\infty}^{\infty} f_{X_1, X_n}(x, r + x)\, dx.$$

Em outras palavras

$$f_R(r) = \begin{cases} n(n-1) \int_{-\infty}^{\infty} f(x) f(r+x)(F(r+x) - F(x))^{n-2}\, dx, & r > 0 \\ 0, & r < 0. \end{cases}$$

Pode-se avaliar facilmente estas fórmulas quando $U_1, \ldots, U_n$ são independentes e se distribuem uniformemente em (0,1). Deixamos esta avaliação para o leitor a título de exercício.

Existe uma forma "heurística" bastante útil para derivar as fórmulas acima. Como ilustração, derivaremos novamente a fórmula para f_{X_k}. Seja dx um número positivo pequeno. Então temos a aproximação

$$f_{X_k}(x)\, dx \approx P(x \leqslant X_k \leqslant x + dx).$$

A maneira mais provável de ocorrer o evento $\{x \leqslant X_k \leqslant x + dx\}$ é $k - 1$ dos U_i estarem em $(-\infty, x]$, um dos U_i em $(x, x + dx]$ e $n - k$ dos U_i em $(x + dx, \infty)$ (ver Figura 5). A derivação da distribuição multinomial dada no Capítulo 3 é aplicável e a probabilidade de que o número indicado de U_i esteja nos intervalos apropriados é

$$\begin{aligned} f_{X_k}(x)\, dx &\approx \frac{n!}{(k-1)!\,1!\,(n-k)!} \\ &\quad \times \left(\int_{-\infty}^{x} f(u)\, du\right)^{k-1} \int_{x}^{x+dx} f(u)\, du \left(\int_{x+dx}^{\infty} f(u)\, du\right)^{n-k} \\ &\approx \frac{n!}{(k-1)!\,(n-k)!} f(x)\, dx F^{k-1}(x)(1 - F(x))^{n-k}, \end{aligned}$$

da qual obtemos (38). Não tentaremos dar rigor a este método.

$$\underset{}{\overset{k-1}{\rule{3cm}{0.4pt}}}\ \underset{x\quad x+dx}{\overset{1}{(\ \ ]}}\ \overset{n-k}{\rule{3cm}{0.4pt}}$$

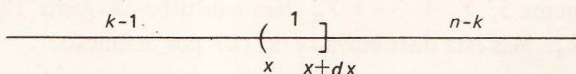



Figura 5

6.6. DISTRIBUIÇÕES AMOSTRAIS*

Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes, cada uma com densidade normal $n(0, \sigma^2)$. Obteremos nesta seção as funções de distribuição de diversas variáveis aleatórias definidas em termos dos X. Além de se constituírem em aplicações do material acima, estas distribuições são de importância fundamental na inferência estatística e serão necessárias no Volume II.

A constante σ^2 é conveniente mas não essencial, uma vez que $X_1/\sigma, \ldots, X_n/\sigma$ são independentes e cada uma tem a densidade normal padrão $n(0, 1)$. Assim podemos sempre tomar $\sigma^2 = 1$ sem perda de generalidade.

De acordo com o Teorema 6, a variável aleatória $X_1 + \cdots + X_n$ tem a densidade normal de parâmetros 0 e $n\sigma^2$. Se dividirmos esta soma por diferentes constantes obtemos formas alternativas deste resultado. Assim

$$\frac{X_1 + \cdots + X_n}{n}$$

distribui-se normalmente com parâmetros 0 e σ^2/n e

$$\frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}}$$

tem densidade normal padrão $n(0, 1)$.

Como X_1/σ tem a densidade normal padrão, segue-se do Exemplo 12 do Capítulo 5 que x_1^2/σ^2 tem a densidade gama $\Gamma(1/2, 1/2)$. Assim, de acordo com o Teorema 5

$$\frac{X_1^2 + \cdots + X_n^2}{\sigma^2}$$

tem a densidade gama $\Gamma(n/2, 1/2)$. Esta densidade gama particular é bastante importante em estatística, onde se diz que a variável aleatória correspondente tem distribuição qui-quadrada (χ^2) com n graus de liberdade que é representada por $\chi^2(n)$. Aplicando o Teorema 5 obtemos o seguinte resultado sobre as distribuições χ^2.

Teorema 7. Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias independentes tais que Y_m tem a distribuição $\chi^2(k_m)$. Então $Y_1 + \cdots + Y_n$ tem a distribuição $\chi^2(k)$, onde $k = k_1 + \cdots + k_n$.

Demonstração. Por hipótese Y_m tem a distribuição gama $\Gamma(k/2, 1/2)$. Assim pelo Teorema 5, $Y_1 + \cdots + Y_n$ têm a distribuição gama $\Gamma(k/2, 1/2)$, onde $k = k_1 + \cdots + k_n$. Mas esta distribuição é $\chi^2(k)$ por definição.

Podemos também aplicar o Teorema 3 para obter a distribuição da razão de duas variáveis aleatórias independentes Y_1, e Y_2 tendo distribuição $\chi^2(k_1)$

e $\chi^2(k_2)$ respectivamente. É tradicional em estatística expressar os resultados em termos das variáveis normalizadas Y_1/k_1 e Y^2/k_2. A distribuição de

$$\frac{Y_1/k_1}{Y_2/k_2}$$

é conhecida como distribuição F com k_1 e k_2 graus de liberdade, sendo representada por $F(k_1, k_2)$.

Teorema 8. Sejam Y_1 e Y_2 variáveis aleatórias independentes tendo distribuições $\chi^2(k_1)$ e $\chi^2(k_2)$. Então a variável aleatória

$$\frac{Y_1/k_1}{Y_2/k_2},$$

que tem a distribuição $F(k_1, k_2)$, tem densidade f dada por $f(x) = 0$ para $x \leqslant 0$ e

$$(39) \qquad f(x) = \frac{(k_1/k_2)\,\Gamma[(k_1 + k_2)/2]\,(k_1x/k_2)^{(k_1/2)-1}}{\Gamma(k_1/2)\,\Gamma(k_2/2)\,[1 + (k_1x/k_2)]^{(k_1+k_2)/2}}, \qquad x > 0.$$

Demonstração. Pelo Teorema 3 a variável aleatória Y_1/Y_2 tem densidade g, onde g é dada por (24) com $\alpha_1 = k_1/2$ e $\alpha_2 = k_2/2$. Assim a densidade f de k_2Y_1/k_1Y_2 é dada por

$$f(x) = \frac{k_1}{k_2}\, g\left(\frac{k_1x}{k_2}\right)$$

e (39) segue-se de (24).

Podemos aplicar este resultado às variáveis aleatórias $X_1, \ldots, X_n$ definidas no início desta seção. Suponha que $1 \leqslant m < n$. Pelo Teorema 4 as variáveis aleatórias

$$\frac{X_1^2 + \cdots + X_m^2}{\sigma^2} \quad \text{e} \quad \frac{X_{m+1}^2 + \cdots + X_n^2}{\sigma^2}$$

são independentes. Como elas têm distribuição $\chi^2(m)$ e $\chi^2(n-m)$ respectivamente, vemos que a variável aleatória

$$\frac{(X_1^2 + \cdots + X_m^2)/m}{(X_{m+1}^2 + \cdots + X_n^2)/(n-m)}$$

tem a distribuição $F(m, n-m)$ e a densidade dada por (39), onde $k_1 = m$ e $k_2 = n - m$. Tabelas das distribuições F são dadas no Volume II.

O caso $m = 1$ é especialmente importante. A variável aleatória

$$\frac{X_1^2}{(X_2^2 + \cdots + X_n^2)/(n-1)}$$

tem distribuição $F(1, n-1)$. Podemos usar este fato para obter a distribuição de

$$Y = \frac{X_1}{\sqrt{(X_2^2 + \cdots + X_n^2)/(n-1)}}.$$

Como X_1 tem uma função de densidade simétrica e é independente da variável aleatória $\sqrt{(X_2^2 + \cdots + X_n^2)/(n-1)}$. Segue-se facilmente do Teorema 2 do Capítulo 5 que Y tem uma função de densidade simétrica f_Y. De acordo com o Exemplo 5 do Capítulo 5, a densidade f_{Y_2} está relacionada com f_Y através de

$$f_{Y^2}(z) = \frac{1}{2\sqrt{z}}(f_Y(-\sqrt{z}) + f_Y(\sqrt{z})), \qquad z > 0.$$

Usando a simétrica de f_Y e fazendo $z = y^2$ vemos que

$$f_Y(y) = |y| f_{Y^2}(y^2).$$

Como Y^2 tem a densidade $F(1, n-1)$ dada por (39) com $k_1 = 1$ e $k_2 = k = n - 1$, obtemos

$$f_Y(y) = \frac{|y|(1/k)\Gamma[(k+1)/2](y^2/k)^{-1/2}}{\Gamma(1/2)\Gamma(k/2)[1 + (y^2/k)]^{(k+1)/2}}.$$

Como $\Gamma(1/2) = \sqrt{\pi}$, esta expressão se reduz a

$$(40) \qquad f_Y(y) = \frac{\Gamma[(k+1)/2][1 + (y^2/k)]^{-(k+1)/2}}{\sqrt{k\pi}\,\Gamma(k/2)}, \qquad -\infty < y < \infty.$$

Diz-se que uma variável aleatória cuja densidade é dada por (40) tem distribuição t com k graus de liberdade. Observamos que a distribuição t com 1 grau de liberdade é a distribuição de Cauchy discutida no Capítulo 4. Tabelas das distribuições t são dadas no Volume II.

A distribuição da variável aleatória

$$Y = \frac{X_1}{\sqrt{(X_2^2 + \cdots + X_n^2)/(n-1)}},$$

que é uma distribuição t com $(n - 1)$ graus de liberdade depende apenas do fato de que

$$\frac{X_1}{\sigma} \quad \text{e} \quad \frac{X_2^2 + \cdots + X_n^2}{\sigma^2}$$

são independentes e têm distribuição $n(0, 1)$ e $\chi^2(n - 1)$, respectivamente. Assim temos o seguinte resultado.

Teorema 9. Sejam X e Y variáveis aleatórias independentes com distribuição $n(0, 1)$ e $\chi^2(k)$, respectivamente. Então

$$\frac{X}{\sqrt{Y/k}}$$

tem distribuição t com k graus de liberdade.

6.7. MUDANÇAS MULTIDIMENSIONAIS DE VARIÁVEIS*

Sejam $X_1, \ldots, X_n$ variáveis aleatórias tendo densidade conjunta f. Sejam $Y_1, \ldots, Y_n$ variáveis aleatórias definidas em termos de X. Nesta seção discutiremos um método para obter a densidade conjunta dos Y em termos de f. Consideraremos principalmente o caso em que se definem os Y como funções lineares de X.

Suponha então que

$$Y_i = \sum_{j=1}^{n} a_{ij} X_j, \qquad i = 1, \ldots, n.$$

Os coeficientes constantes a_{ij} determinam uma matriz $n \times n$

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}.$$

Associada a esta matriz está o seu determinante

$$\det A = \begin{vmatrix} a_{11} & \cdots & a_{11} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}.$$

Se $\det A \neq 0$, existe uma única matriz inversa $B = [\, b_{ij} \,]$ tal que $BA = I$ ou equivalentemente

$$(41) \qquad \sum_{k=1}^{n} b_{ik} a_{kj} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Pode-se obter as constantes b_{ij} resolvendo para cada i o sistema (41) de n equações a n incógnitas $b_{i_1}, \ldots, b_{i_n}$. Alternativamente, as constantes b_{ij} podem ser unicamente definidas exigindo que as equações

$$y_i = \sum_{j=1}^{n} a_{ij}x_j, \qquad i = 1, \ldots, n,$$

tenham soluções

$$x_i = \sum_{j=1}^{n} b_{ij}y_j, \qquad i = 1, \ldots, n. \tag{42}$$

Teorema 10. Sejam $X_1, \ldots, X_n$ variáveis aleatórias tendo densidade conjunta f e suponha que as variáveis aleatórias $Y_1, \ldots, Y_n$ são definidas por

$$Y_i = \sum_{j=1}^{n} a_{ij}X_j, \qquad i = 1, \ldots, n,$$

onde a matriz $A = [\, a_{ij} \,]$ tem determinante não-nulo $\det A$. Então $Y_1, \ldots, Y_n$ têm a densidade conjunta $f_{Y_1}, \ldots, f_{Y_n}$ dada por

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \frac{1}{|\det A|} f(x_1, \ldots, x_n), \tag{43}$$

onde os x são definidos em termos de y por (42) ou como a solução única das equações $y_i = \sum_{j=1}^{n} a_{ij}x_j$.

Este teorema, que não será demonstrado aqui, é equivalente ao teorema demonstrado em cursos avançados de cálculo em um contexto mais geral envolvendo "Jacobianos". Do resultado geral demonstrado em cálculo avançado, podemos estender o teorema acima a mudanças não-lineares de variáveis. Descreveremos brevemente esta extensão, embora ela não seja necessária mais tarde.

Suponha que os Y são definidos em termos de X por

$$Y_i = g_i(X_1, \ldots, X_n), \qquad i = 1, \ldots, n.$$

Considere as equações correspondentes

$$y_i = g_i(x_1, \ldots, x_n), \qquad i = 1, \ldots, n. \tag{44}$$

Suponha que estas equações definem unicamente os x em termos de y, que as derivadas parciais $\partial y_i/\partial x_j$ existem e são contínuas e que o Jacobiano

$$J(x_1, \ldots, x_n) = \begin{vmatrix} \dfrac{\partial y_1}{\partial x_1} & \cdots & \dfrac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial y_n}{\partial x_1} & \cdots & \dfrac{\partial y_n}{\partial x_n} \end{vmatrix}$$

é sempre não-nulo. Então as variáveis aleatórias $Y_1, \ldots, Y_n$ são contínuas e têm uma densidade conjunta dada por

$$(45) \qquad f_{Y_1,\ldots,Y_n}(y_1, \ldots y_n) = \frac{1}{|J(x_1, \ldots, x_n)|} f(x_1, \ldots, x_n),$$

onde os x são definidos implicitamente em termos de y por (44). Pode-se estender ainda mais esta fórmula de mudança de variável, exigindo que as funções g_i sejam definidas, apenas em algum subconjunto aberto S de R^n tal que

$$P((X_1, \ldots, X_n) \in S) = 1.$$

No caso especial em que $y_i = \sum_{j=1}^n a_{ij}x_j$, vemos que $\partial y_i/\partial x_j = a_{ij}$ e $J(x_1, \ldots x_n)$ é simplesmente a constante $\det A = \det | a_{ij} |$. Assim, torna-se claro que (45) se reduz a (43) no caso linear.

Exemplo 13. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes, cada uma com densidade exponencial de parâmetro λ. Defina $Y_1, \ldots, Y_n$ por $Y_i = X_1 + \cdots + X_i$, $1 \leqslant i \leqslant n$. Obtenha a densidade conjunta de $Y_1, \ldots, Y_n$.

A matriz $[a_{ij}]$ é

$$\begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 1 & 1 & 0 & & & \vdots \\ \vdots & & & & & 0 \\ 1 & 1 & \cdot & \cdot & \cdot & 1 \end{bmatrix}$$

Seu determinante é claramente 1. As equações

$$y_i = x_1 + \cdots + x_i, \qquad i = 1, \ldots, n,$$

têm a solução

$$x_1 = y_1,$$
$$x_i = y_i - y_{i-1}, \qquad i = 2, \ldots, n.$$

A densidade conjunta de $X_1, \ldots, X_n$ é dada por

$$(46) \qquad f(x_1, \ldots, x_n) = \begin{cases} \lambda^n e^{-\lambda(x_1 + \cdots + x_n)}, & x_1, \ldots, x_n > 0, \\ 0, & \text{para outros valores de } x_1, \ldots, x_n \end{cases}$$

Assim a densidade conjunta $f_{Y_1,\ldots,Y_n}$ é dada por

$$(47) \qquad f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \begin{cases} \lambda^n e^{-\lambda y_n}, & 0 < y_1 < \cdots < y_n, \\ 0, & \text{para outros valores de } y_1, \ldots, y_n \end{cases}$$

Naturalmente podemos aplicar o teorema em sentido inverso. Assim se $Y_1, \ldots, Y_n$ têm a densidade conjunta dada por (47), e as variáveis aleatórias $X_1, \ldots, X_n$ são definidas por $X_1 = Y_1$ e $X_i = Y_i - Y_{i-1}$ para $2 \leqslant i \leqslant n$, então os X têm a densidade conjunta dada por (46). Em outras palavras, $X_1, \ldots, X_n$ são independentes, cada uma com distribuição exponencial de parâmetro λ. Usaremos este resultado no Capítulo 9 na discussão de processos de Poisson.

1. Sejam X e Y variáveis aleatórias contínuas tendo função de densidade conjunta f. Obtenha a função de distribuição conjunta e a função de densidade conjunta das variáveis aleatórias $W = a + bX$ e $Z = c + dY$, onde $b > 0$ e $d > 0$. Mostre que se X e Y são independentes, então W e Z são independentes.
2. Sejam X e Y variáveis aleatórias contínuas tendo função de distribuição conjunta F e função de densidade conjunta f. Obtenha a função de distribuição conjunta e a função de densidade conjunta das variáveis aleatórias $W = X^2$ e $Z = Y^2$. Mostre que se X e Y são independentes, então W e Z são independentes.
3. Sejam X e Y variáveis aleatórias independentes, cada uma com distribuição uniforme em $(0,1)$. Obtenha

   (a) $P(|X - Y| \leqslant 0{,}5)$,

   (b) $P\left(\left|\dfrac{X}{Y} - 1\right| \leq 0{,}5\right)$,

   (c) $P(Y \geqslant X \mid Y \geqslant 1/2)$.

4. Sejam X e Y variáveis aleatórias independentes, cada uma com distribuição normal $n(0, \sigma^2)$. Obtenha $P(X^2 + Y^2 \leqslant 1)$. *Sugestão:* usar coordenadas polares.
5. Suponha que X e Y têm uma densidade conjunta f uniforme no interior do triângulo com vértices em $(0, 0)$, $(2, 0)$ e $(1, 2)$. Obtenha $P(X \leqslant 1$ e $Y \leqslant 1)$.
6. Suponha que os tempos que dois estudantes levam para resolver um problema são independentes e se distribuem exponencialmente com parâmetro λ. Determine a probabilidade de que o primeiro estudante necessite pelo menos do dobro do tempo gasto pelo segundo estudante para resolver o problema.
7. Sejam X e Y variáveis aleatórias contínuas tendo função de densidade conjunta dada por $f(x, y) = \lambda^2 e^{-\lambda y}$, $0 \leqslant x \leqslant y$, e $f(x, y) = 0$ para outros valores de x e y. Obtenha as densidades marginais de X e Y. Obtenha a função de distribuição conjunta de X e Y.
8. Seja $f(x, y) = c(y - x)^{\alpha}$, $0 \leqslant x < y \leqslant 1$, e $f(x, y) = 0$ para outros valores de x e y.

   (a) Para que valores de α se pode escolher c para fazer de f uma função de densidade?

   (b) Como se deve escolher c (quando possível) para fazer de f uma densidade?

   (c) Obtenha as densidades marginais de f.

9. Seja $f(x, y) = ce^{-(x^2 - xy + 4y^2)/2}$, $-\infty < x$, $y < \infty$. Como se deve escolher c para fazer de f uma densidade? Obtenha as densidades marginais de f.

10. Sejam X e Y variáveis aleatórias contínuas independentes tendo densidade conjunta f. Derive uma fórmula para a densidade de $Z = Y - X$.

11. Sejam X e Y variáveis aleatórias contínuas independentes tendo as densidades marginais especificadas abaixo. Obtenha a densidade de $Z = X + Y$.

    (a) X e Y distribuem-se exponencialmente com parâmetros λ_1 e λ_2, onde $\lambda_1 \neq \lambda_2$.

    (b) X é uniforme em $(0,1)$ e Y distribui-se exponencialmente com parâmetro λ.

12. Suponha que X e Y têm a densidade conjunta f dada no Exercício 8. Obtenha a densidade de $Z = X + Y$.

13. Suponha que X e Y são independentes e distribuem-se uniformemente em (a, b). Obtenha a densidade de $Z = |Y - X|$.

14. Sejam X e Y variáveis aleatórias contínuas tendo densidade conjunta f. Derive a fórmula para a densidade de $Z = aX + bY$, onde $b \neq 0$.

15. Seja f uma densidade Beta de parâmetros $\alpha_1 > 1$ e $\alpha_2 > 1$. Em que ponto f assume o seu valor máximo?

16. Sejam X e Y variáveis aleatórias independentes, tendo as densidades normais $n(\mu_1, \sigma_1^2)$ e $n(\mu_2, \sigma_2^2)$, respectivamente. Obtenha a densidade de $Z = Y - X$.

17. Suponha que se escolhe aleatoriamente um ponto no plano de tal forma que suas coordenadas se distribuem independentemente segundo a densidade normal $n(0, \sigma^2)$. Obtenha a função de densidade da variável aleatória R que representa a distância do ponto escolhido à origem. (Esta densidade ocorre em engenharia elétrica, sendo conhecida como densidade de *Rayleigh*).

18. Sejam X e Y variáveis aleatórias contínuas tendo densidade conjunta f. Derive a fórmula para a densidade de $Z = XY$.

19. Sejam X e Y variáveis aleatórias independentes, cada uma com a densidade normal $n(0, \sigma^2)$. Mostre que tanto Y/X como $Y/|X|$ têm a densidade de Cauchy.

20. Sejam X e Y definidos como no Exercício 19. Obtenha a densidade de $Z = |Y|/|X|$.

21. Sejam X e Y duas variáveis aleatórias independentes, cada uma com distribuição exponencial de parâmetro λ. Obtenha a densidade de $Z = Y/X$.

22. Sejam X e Y duas variáveis aleatórias independentes tendo as densidades gama $\Gamma(\alpha_1, \lambda)$ e $\Gamma(\alpha_2, \lambda)$, respectivamente. Obtenha a densidade de $Z = X/(X + Y)$. *Sugestão:* expresse Z em têrmos de Y/X.

23. Suponha que X e Y têm a densidade conjunta f indicada abaixo. Obtenha a densidade condicional $f_{Y|X}$ em cada caso.

    (a) f como no Exercício 7,

    (b) f como no Exercício 8,

    (c) f como no Exercício 9.

24. Suponha que X e Y se distribuem como as variáveis definidas no Exemplo 7. Determine $P(Y \leqslant 2 \mid X = 1)$.

25. Mostre que a densidade marginal f_Y no Exemplo 9 é binomial negativa com parâmetros α e $p = \beta/(\beta + 1)$. *Sugestão:* use a fórmula (36) do Capítulo 5.

26. Seja Y uma variável aleatória discreta tendo a distribuição binomial de parâmetros n e p. Suponha que p se comporta como uma variável aleatória π tendo a densidade Beta de parâmetros α_1 e α_2. Obtenha a densidade condicional de π, dado $Y = y$.

27. Suponha que Y se distribui exponencialmente com parâmetro λ. Suponha que λ se comporta como uma variável aleatória Λ tendo a densidade gama $\Gamma(\alpha, \beta)$. Obtenha a densidade marginal de Y e a densidade condicional de Λ dado $Y = y$.

28. Suponha que X_1, X_2, e X_3 representam os componentes da velocidade de uma molécula de gás. Suponha que X_1, X_2 e X_3 são independentes, cada uma com densidade normal $n(0, \sigma^2)$. Diz-se em física que a magnitude da velocidade $Y = (X_1^2 + X_2^2 + X_3^2)^{1/2}$ tem uma distribuição de *Maxwell*. Obtenha f_Y.

29. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes tendo uma densidade normal comum. Mostre que existem constantes A_n e B_n tais que

$$\frac{X_1 + \cdots + X_n - A_n}{B_n}$$

têm a mesma densidade de X_1.

30. Sejam X_1, X_2 e X_3 variáveis aleatórias independentes, individualmente uniformes em $(0,1)$. Obtenha a densidade da variável aleatória $Y = X_1 + X_2 + X_3$. Determine $P(X_1 + X_2 + X_3 \leqslant 2)$.

31. Suponha que se escolhe X_1 uniformemente em $(0, 1)$, X_2 uniformemente em $(0, X_1)$ e que se escolhe X_3 uniformemente em $(0, X_2)$. Obtenha a densidade conjunta de X_1, X_2 e X_3 e a densidade marginal de X_3.

32. Sejam $U_1, \ldots, U_n$ variáveis aleatórias independentes, cada uma com distribuição uniforme em $(0, 1)$. Sejam X_k, $k = 1, \ldots, n$, e R definidos como na Seção 6.5.

    (a) Obtenha a densidade conjunta de X_1 e X_n.

    (b) Obtenha a densidade de R.

    (c) Obtenha a densidade de X_k.

33. Sejam $U_1, \ldots, U_n$ variáveis aleatórias independentes, cada uma com densidade exponencial de parâmetro λ. Obtenha a densidade de $X_1 = \min(U_1, \ldots, U_n)$.

34. Obtenha uma fórmula para a densidade $\chi^2(n)$.

35. Sejam X e Y variáveis aleatórias independentes tendo as densidades $\chi^2(m)$ e $\chi^2(n)$, respectivamente. Obtenha a densidade de $Z = X/(X + Y)$. *Sugestão:* use a resposta do Exercício 22.

36. Sejam X e Y variáveis aleatórias independentes, cada uma com a densidade normal padrão. Obtenha a densidade conjunta de $aX + bX$ e $bX - aY$, onde $a^2 + b^2 > 0$. Use este resultado para dar uma outra demonstração ao Teorema 2.

37. Sejam X e Y variáveis aleatórias independentes com densidade comum f. Obtenha a densidade conjunta de X e $Z = X + Y$.

38. Sejam X e Y variáveis aleatórias independentes, cada uma com a densidade exponencial de parâmetro λ. Determine a densidade condicional de X dado $Z = X + Y = z$. *Sugestão:* use o resultado do Exercício 37.

39. Resolva o Exercício 38 para o caso em que X e Y se distribuam uniformemente em $(0, c)$.

40. Suponha que U e V são variáveis aleatórias independentes, cada uma com densidade normal padrão. Seja $Z = \rho U + \sqrt{1 - \rho^2}\, V$, onde $-1 < \rho < 1$.
    (a) Obtenha a densidade de Z.
    (b) Obtenha a densidade conjunta de U e Z.
    (c) Obtenha a densidade conjunta de $X = \mu_1 + \sigma_1 U$ e $Y = \mu_2 + \sigma_2 Z$, onde $\sigma_1 > 0$ e $\sigma_2 > 0$. Esta densidade é conhecida como *densidade normal bidimensional.*
    (d) Obtenha a densidade condicional de Y dado $X = x$.

41. Sejam X e Y variáveis aleatórias contínuas positivas com uma densidade conjunta f. Sejam $W = Y/X$ e $Z = X + Y$. Obtenha a densidade conjunta de W e Z em termos de f. *Sugestão:* use a Equação (45).

42. Sejam X e Y variáveis aleatórias independentes tendo as densidades gama $\Gamma(\alpha_1, \lambda)$ e $\Gamma(\alpha_2, \lambda)$, respectivamente. Use o Exercício 41 para mostrar que Y/X e $X + Y$ são variáveis aleatórias independentes.

43. Sejam R e Θ variáveis aleatórias independentes tais que R tem a densidade de Rayleigh

$$f_R(r) = \begin{cases} \sigma^{-2} r\, e^{-r^2/2\sigma^2}, & r \geq 0, \\ 0, & r < 0, \end{cases}$$

e Θ se distribui uniformemente em $(-\pi, \pi)$. Mostre que $X = R \cos \Theta$ e $Y = R \operatorname{sen} \Theta$ são variáveis aleatórias independentes e que cada uma tem a densidade normal $n(0, \sigma^2)$. *Sugestão:* use a equação (45).

7

EXPECTÂNCIAS E O TEOREMA DO LIMITE CENTRAL

Nas primeiras quatro seções deste capítulo estudamos a definição e propriedades das expectâncias de variáveis aleatórias que não são necessariamente discretas. Na Seção 7.5. discutimos o Teorema do Limite Central. Este teorema, um dos mais importantes na teoria da probabilidade, justifica a aproximação de muitas funções de distribuição pela função normal apropriada.

7.1. EXPECTÂNCIAS DE VARIÁVEIS ALEATÓRIAS CONTÍNUAS

Lembremos a definição e expectância de uma variável aleatória discreta X de densidade f (Capítulo 4). Dizemos que X tem expectância finita se $\Sigma_x |x| f(x) < \infty$, e neste caso definimos sua expectância EX, como

$$EX = \sum_x xf(x).$$

A maneira mais fácil de definir expectâncias de variáveis aleatórias contínuas que possuem densidades é em analogia com o caso discreto.

Definição 1. Seja X uma variável aleatória contínua de densidade f. Dizemos que X tem expectância finita se

$$\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty,$$

e neste caso definimos sua expectância por meio de

$$EX = \int_{-\infty}^{\infty} xf(x)\, dx.$$

Usando esta definição, podemos calcular facilmente as expectâncias de variáveis aleatórias contínuas tendo as diversas densidades discutidas nos Capítulos 5 e 6.

Exemplo 1. Suponha que X se distribui uniformemente em (a, b). Então

$$EX = \int_a^b x\left(\frac{1}{b-a}\right) dx = \left(\frac{1}{b-a}\right)\frac{x^2}{2}\bigg|_a^b = \frac{a+b}{2}.$$

Exemplo 2. Suponha que X tem a densidade gama $\Gamma(\alpha, \lambda)$. Então

$$\begin{aligned} EX &= \int_0^\infty x \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\lambda x}\, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \\ &= \frac{\alpha}{\lambda}, \end{aligned}$$

onde usamos as fórmulas (34) e (36) do Capítulo 5. Fazendo $\alpha = 1$ vemos que se X tem uma densidade exponencial de parâmetro λ, então $EX = \lambda^{-1}$.

Exemplo 3. Suponha que X tem a densidade de Cauchy f dada por

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad -\infty < x < \infty.$$

Então X não tem expectância finita, pois

$$\begin{aligned} \int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)}\, dx &= \frac{2}{\pi} \int_0^\infty \frac{x}{1+x^2}\, dx \\ &= \frac{2}{\pi} \lim_{c\to\infty} \int_0^c \frac{x}{1+x^2}\, dx \\ &= \frac{1}{\pi} \lim_{c\to\infty} \log(1+x^2)\Big|_0^c \\ &= \infty. \end{aligned}$$

7.2. UMA DEFINIÇÃO GERAL DE EXPECTÂNCIA

A definição de expectância dada na Seção 7.1 é certamente apropriada do ponto de vista computacional para o caso de variáveis aleatórias contínuas que possuem densidades. Entretanto, para definir expectância em geral é melhor estender a noção de expectância diretamente do caso discreto ao caso geral. Os detalhes precisos requerem conhecimentos adicionais da teoria da medida e integração. Vamos supor em nossa discussão que todas as variáveis aleatórias em consideração são definidas em um espaço fixo de probabilidade $(\Omega, \mathscr{A}, P)$.

Sejam X e Y variáveis aleatórias discretas tais que $P(|X - Y| \leqslant \epsilon) = 1$ para algum $\epsilon > 0$. Segue-se dos Teoremas 2 (iii) e 3 do Capítulo 4 que se Y tem expectância finita, X também a tem e $|EX - EY| \leqslant \epsilon$. Segue-se também que se Y não tem expectância finita, X também não tem. Quando se define expectância em geral, estas propriedades devem continuar sendo válidas.

Suponhamos que este é o caso e seja X uma variável aleatória qualquer. Suponha que desejamos calcular EX com erro não superior a ϵ, para algum $\epsilon > 0$. Tudo que precisamos fazer é obter uma variável aleatória discreta Y tal que $P(|X - Y| \leqslant \epsilon) = 1$ e determinar EY de acordo com os métodos do Capítulo 4.

É fácil obter tais aproximações para X. Seja X_ϵ a variável aleatória discreta definida por

(1) $$X_\epsilon = \epsilon k \quad \text{se} \quad \epsilon k \leqslant X < \epsilon(k+1) \text{ para } k \text{ inteiro.}$$

Pode-se também definir esta variável aleatória em termos da função de maior inteiro [ ] como $X_\epsilon = \epsilon[X/\epsilon]$. Se $\epsilon = 10^{-n}$ para algum inteiro não-negativo n, $X_\epsilon(\omega)$ pode ser obtido de $X(\omega)$ escrevendo $X(\omega)$ em forma decimal e desprezando todos os dígitos que estiverem a n ou mais casas além do ponto decimal. Segue-se imediatamente de (1) que

$$X(\omega) - \epsilon < X_\epsilon(\omega) \leqslant X(\omega), \qquad \omega \in \Omega,$$

e portanto $P(|X - X_\epsilon| \leqslant \epsilon) = 1$. A função de densidade de X_ϵ é dada por

$$f_{X_\varepsilon}(x) = \begin{cases} P(\varepsilon k \leq X < \varepsilon(k+1)) & \text{se } x = \epsilon k \text{ para } k \text{ inteiro} \\ 0 & \text{para outros valores de } x. \end{cases}$$

A variável aleatória X_ϵ tem expectância finita se e somente se

$$\sum_x |x| f_{X_\varepsilon}(x) = \sum_k |\varepsilon k| P(\varepsilon k \leq X < \varepsilon(k+1)) < \infty,$$

neste caso

$$EX_\varepsilon = \sum_k \varepsilon k P(\varepsilon k \leq X < \varepsilon(k+1)).$$

Pode-se escrever estas expressões em termos de F_X. Para

$$P(\epsilon k \leqslant X < \epsilon(k+1)) = P(X < \epsilon(k+1)) - P(X < \epsilon k)$$

e pela Equação (5) do Capítulo 5, $P(X < x) = F(x-)$ se verifica para todo x. O teorema seguinte, que enunciamos sem demonstrar, será usado para dar uma definição geral de expectância.

Teorema 1. Seja X uma variável aleatória e suponha que X_ϵ, $\epsilon > 0$, seja definido por (1). Se X_ϵ tem expectância finita para algum $\epsilon > 0$, então X_ϵ tem expectância finita para todo $\epsilon > 0$ e

$$\lim_{\varepsilon \to \infty} EX_\varepsilon$$

existe e é finito.

Este teorema e a nossa discussão precedente sugerem a seguinte definção geral de expectância.

Definição 2. Seja X uma variável aleatória e suponha que X_ϵ, $\epsilon > 0$, seja definido por (1). Se X_ϵ tem expectância finita para algum $\epsilon > 0$, dizemos que X tem expectância finita e definimos sua expectância EX através de

$$EX = \lim_{\varepsilon \to 0} EX_\varepsilon .$$

No caso contrário dizemos que X não tem expectância finita.

Da discussão que precede o Teorema 1 segue-se que a definição de EX pode ser dada em termos da função de distribuição de X e que se duas variáveis aleatórias têm a mesma função de distribuição, suas expectâncias são iguais (ou ambas não são finitas). Usando técnicas da teoria da medida e integração, pode-se mostrar que a Definição 2 conduz aos mesmos valores que as definições anteriores para os casos em que X é discreto ou quando X é uma variável aleatória contínua que possui uma densidade. Existe um análogo do Teorema 1 do Capítulo 4 que enunciamos sem demonstração. Neste teorema, φ pode ser qualquer função do tipo considerado em cálculo.

Teorema 2. Sejam $X_1, \ldots, X_n$ variáveis aleatórias contínuas tendo densidade conjunta f e seja Z uma variável aleatória definida em termos de $X_1, \ldots, X_n$ através de $Z = \varphi(X_1, \ldots, X_n)$. Então Z tem expectância finita se e somente se

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |\varphi(x_1, \ldots, x_n)| f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n < \infty,$$

e neste caso

$$EZ = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(x_1, \ldots, x_n) f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n .$$

Podemos mostrar que as propriedades básicas da expectância demonstradas no Capítulo 4 para variáveis aleatórias discretas são válidas em geral. Em particular os Teoremas 2, 3 e 4 do Capítulo 4 são válidos e serão usados livremente.

Como no caso discreto, nos referimos, às vezes, a EX como *média* de X. As definições de momentos, momento central, variância, desvio padrão, covariância e correlação dadas no Capítulo 4 para variáveis aleatórias discretas dependem somente da noção de expectância e estendem-se imediatamente para o caso geral.

Em geral, como no caso discreto, se X tem um momento de ordem r, então X tem um momento de ordem k para todo $k \leq r$. Os Teoremas 6 e 7 do Capítulo 4 são também válidos em geral. O leitor deve rever os teoremas e definições mencionadas no Capítulo 4 antes de prosseguir para a seção seguinte.

7.3. MOMENTOS DE VARIÁVEIS ALEATÓRIAS CONTÍNUAS

Seja X uma variável aleatória contínua tendo densidade f e média μ. Se X tem um m-ésimo momento finito, então pelo Teorema 2

$$EX^m = \int_{-\infty}^{\infty} x^m f(x)\, dx$$

e

$$E(X - \mu)^m = \int_{-\infty}^{\infty} (x - \mu)^m f(x)\, dx.$$

Em particular, se X tem segundo momento finito, sua variância σ^2 é dada por

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx.$$

Observe que $\sigma^2 > 0$. Pois se $\sigma^2 = 0$, teríamos pelo argumento da Seção 4.3 que $P(X = \mu) = 1$, o que contradiz a hipótese de que X é uma variável aleatória contínua.

Exemplo 4. Suponha que X tem a densidade gama $\Gamma(\alpha, \lambda)$. Obtenha os momentos e a variância de X.

O m-ésimo momento de X é dado por

$$\begin{aligned} EX^m &= \int_0^{\infty} x^m \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\, dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{m+\alpha-1} e^{-\lambda x}\, dx, \end{aligned}$$

de modo que pelas fórmulas (34) e (36) do Capítulo 5

$$\begin{aligned} EX^m &= \frac{\lambda^\alpha \Gamma(m + \alpha)}{\lambda^{m+\alpha}\Gamma(\alpha)} \\ &= \frac{\alpha(\alpha + 1)\cdots(\alpha + m - 1)}{\lambda^m}. \end{aligned} \tag{2}$$

A variância de X é dada por

$$\sigma^2 = EX^2 - (EX)^2 = \frac{\alpha(\alpha + 1)}{\lambda^2} - \left(\frac{\alpha}{\lambda}\right)^2 = \frac{\alpha}{\lambda^2}.$$

Fazendo $\alpha = 1$, vemos que se X tem densidade exponencial de parâmetro λ, então $EX^m = m!\, \lambda^{-m}$ e X tem variância λ^{-2}. Como um segundo caso especial, suponha que X tem a distribuição $\chi^2(n)$ que, de acordo com a Seção 6.6, é a distribuição $\Gamma(n/2, 1/2)$. Então

$$EX = \frac{n/2}{1/2} = n \quad \text{e} \quad \operatorname{Var} X = \frac{n/2}{(1/2)^2} = 2n.$$

Freqüentemente podemos tirar partido da simetria na determinação de momentos. Por exemplo, se X tem densidade simétrica, se EX^m existe e se m é um número inteiro positivo ímpar, então $EX^m = 0$. Para verificar este resultado, observe que pelo Teorema 2 do Capítulo 5, X e $-X$ têm a mesma função de distribuição. Assim X^m e $(-X)^m = -X^m$ tem a mesma função de distribuição e conseqüentemente a mesma expectância. Em outras palavras $EX^m = E(-X^m) = -EX^m$, o que implica em $EX^m = 0$.

Exemplo 5. Suponha que X tem a densidade normal $n(\mu, \sigma^2)$. Obtenha a média e os momentos centrais de X.

A variável aleatória $X - \mu$ tem a densidade normal $n(0, \sigma^2)$, que é simétrica. Assim $E(X - \mu)^m = 0$ para todo número inteiro positivo ímpar m. Em particular $E(X - \mu) = 0$, o que mostra que o parâmetro μ da densidade normal $n(\mu, \sigma^2)$ é simplesmente a média da densidade. Segue-se então que todos os momentos centrais de ordem ímpar de X são iguais a zero. Para determinar os momentos centrais de ordem par, lembramos da Seção 5.3.3 que $Y = (X - \mu)^2$ tem a densidade gama $\Gamma(1/2, 1/2\sigma^2)$. Como $E(X - \mu)^m = EY^{m/2}$, para m par, segue-se do Exemplo 4 que

$$E(X - \mu)^m = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\left(\frac{1}{2\sigma^2}\right)^{m/2}\Gamma\left(\frac{1}{2}\right)}$$

$$= \frac{\frac{1}{2}\cdot\frac{3}{2}\cdots\left(\frac{m-1}{2}\right)}{\left(\frac{1}{2\sigma^2}\right)^{m/2}}$$

$$= \sigma^m 1 \cdot 3 \cdots (m - 1).$$

Usando as fórmulas (35) e (38) do Capítulo 5, obtemos a fórmula alternativa

$$(3) \qquad E(X - \mu)^m = \frac{m!}{2^{m/2}\left(\frac{m}{2}\right)!}\sigma^m.$$

Em particular σ^2 representa as variâncias de X e $E(X - \mu)^4 = 3\sigma^4$.

Sejam X e Y variáveis aleatórias contínuas tendo densidade conjunta f, médias μ_X e μ_Y, e segundos momentos finitos. Sua covariância é dada por

$$(4) \qquad E(X - \mu_X)(Y - \mu_Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_Y)(y - \mu_Y)f(x, y)\, dx\, dy.$$

Exemplo 6. Suponha que X e Y têm a densidade conjunta f do Exemplo 2 do Capítulo 6. Obtenha a correlação entre X e Y. De acordo com o Exemplo 2 do Capítulo 6.

$$f(x, y) = \frac{\sqrt{3}}{4\pi} e^{-[(x^2 - xy + y^2)/2]}$$

$$= \frac{\sqrt{3}}{4\pi} e^{-3x^2/8} e^{-[(y-x/2)^2/2]}.$$

Vimos naquele exemplo que tanto X como Y têm a densidade normal $n(0, 4/3)$. Assim $\mu_X = \mu_Y = 0$ e $\text{Var } X = \text{Var } Y = 4/3$. Da Equação (4) e da segunda expressão de f temos

$$EXY = \frac{\sqrt{3}}{2\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-(3x^2/8)}\, dx \int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-[(y-x/2)^2/2]}\, dy.$$

Mas

$$\int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi}} e^{-[(y-x/2)^2/2]}\, dy = \int_{-\infty}^{\infty} \left(u + \frac{x}{2}\right) \frac{1}{\sqrt{2\pi}} e^{-(u^2/2)}\, du = \frac{x}{2},$$

de modo que

$$EXY = \frac{1}{2\left(\frac{2}{\sqrt{3}}\right)\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-(3x^2/8)}\, dx$$

$$= 1/2 \int_{-\infty}^{\infty} x^2 n(x; 0, 4/3)\, dx$$

$$= 1/2 \cdot 4/3 = 2/3.$$

A correlação ρ entre X e Y é dada por

$$\rho = \frac{EXY}{\sqrt{\text{Var } X}\,\sqrt{\text{Var } Y}} = \frac{2/3}{\sqrt{4/3}\,\sqrt{4/3}} = \frac{1}{2}.$$

Exemplo 7. Sejam $U_1, \ldots, U_n$ variáveis aleatórias independentes, cada uma com distribuição uniforme em $(0, 1)$ e seja

$$X = \min(U_1, \ldots, U_n)$$

e

$$Y = \max(U_1, \ldots, U_n).$$

Obtenha os momentos de X e Y e a correlação entre X e Y.

Estas variáveis aleatórias foram estudadas na Seção 6.5 (onde foram representadas por X_1 e X_n). Especializando os resultados daquela seção para U_i que se distribuem uniformemente, concluímos que X e Y têm a densidade conjunta f dada por

$$(5) \qquad f(x, y) = \begin{cases} n(n-1)(y-x)^{n-2}, & 0 \le x \le y \le 1, \\ 0, & \text{outros valores de } x \text{ e } y. \end{cases}$$

Os leitores que omitiram a Seção 6.5 podem pensar no presente problema como o de obter a correlação entre as variáveis aleatórias X e Y cuja densidade conjunta é dada por (5).

O m-ésimo momento de X é dado por

$$\begin{aligned} EX^m &= n(n-1) \int_0^1 x^m \, dx \int_x^1 (y-x)^{n-2} \, dy \\ &= n(n-1) \int_0^1 x^m \, dx \left. \frac{(y-x)^{n-1}}{n-1} \right|_{y=x}^{y=1} \\ &= n \int_0^1 x^m (1-x)^{n-1} \, dx. \end{aligned}$$

A integral definida que aparece nesta expressão é uma integral Beta e foi determinada na fórmula (19) do Capítulo 6. Daquela fórmula obtemos

$$EX^m = \frac{n\Gamma(m+1)\Gamma(n)}{\Gamma(m+n+1)} = \frac{m!\, n!}{(m+n)!}.$$

Em particular $EX = 1/(n+1)$ e $EX^2 = 2/(n+1)(n+2)$. Segue-se que

$$\operatorname{Var} X = (EX^2) - (EX)^2 = \frac{n}{(n+1)^2(n+2)}.$$

O m-ésimo momento de Y é dado por

$$\begin{aligned} EY^m &= n(n-1) \int_0^1 y^m \, dy \int_0^y (y-x)^{n-2} \, dx \\ &= n(n-1) \int_0^1 y^m \, dy \left. \frac{(y-x)^{n-1}(-1)}{n-1} \right|_{x=0}^{x=y} \\ &= n \int_0^1 y^{m+n-1} \, dy \\ &= \frac{n}{m+n}. \end{aligned}$$

Assim $EY = n/(n+1)$ e

$$\operatorname{Var} Y = \frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2 = \frac{n}{(n+1)^2(n+2)}.$$

Alternativamente, estes resultados podem ser obtidos das densidades marginais de X e Y.

Para obter a covariância de X e Y começamos com

$$EXY = n(n-1)\int_0^1 y\,dy \int_0^y x(y-x)^{n-2}\,dx.$$

Como

$$x(y-x)^{n-2} = y(y-x)^{n-2} - (y-x)^{n-1},$$

obtemos

$$\begin{aligned} EXY &= n(n-1)\int_0^1 y^2\,dy \int_0^y (y-x)^{n-2}\,dx \\ &\quad - n(n-1)\int_0^1 y\,dy \int_0^y (y-x)^{n-1}\,dx \\ &= n(n-1)\int_0^1 y^2\,dy \left.\frac{(y-x)^{n-1}(-1)}{n-1}\right|_{x=0}^{x=y} \\ &\quad - n(n-1)\int_0^1 y\,dy \left.\frac{(y-x)^{n}(-1)}{n}\right|_{x=0}^{x=y} \\ &= n\int_0^1 y^{n+1}\,dy - (n-1)\int_0^1 y^{n+1}\,dy \\ &= \frac{1}{n+2}. \end{aligned}$$

Conseqüentemente

$$\begin{aligned} \operatorname{Cov}(X, Y) &= EXY - EXEY \\ &= \frac{1}{n+2} - \frac{n}{(n+1)^2} \\ &= \frac{1}{(n+1)^2(n+2)}. \end{aligned}$$

Finalmente obtemos para a correlação entre X e Y

$$\begin{aligned} \rho &= \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var} X \operatorname{Var} Y}} \\ &= \frac{1}{(n+1)^2(n+2)} \bigg/ \frac{n}{(n+1)^2(n+2)} \\ &= \frac{1}{n}. \end{aligned}$$

7.4. EXPECTÂNCIA CONDICIONAL

Sejam X e Y variáveis contínuas tendo densidade conjunta f e suponha que Y tem expectância finita. Na Seção 6.3 definimos a densidade condicional de Y dado $X = x$ através de

$$f_{Y|X}(y \mid x) = \begin{cases} \dfrac{f(x, y)}{f_X(x)}, & 0 < f_X(x) < \infty, \\ 0, & \text{para outros valores de } x \text{ e } y. \end{cases}$$

A função $f_{Y|X}(y \mid x)$, $-\infty < y < \infty$, é uma função de densidade para cada x tal que $0 < f_X(x) < \infty$ de acordo com a Definição 5 do Capítulo 5. Assim podemos pensar em vários momentos desta densidade. A sua média chama-se *expectância condicional* de Y dado $X = x$, sendo representada por

$E[\, Y \mid X = x\,]$ ou $E[\, Y \mid x\,]$. Assim

$$E[Y \mid X = x] = \int_{-\infty}^{\infty} yf(y \mid x)\, dy \tag{6}$$

$$= \frac{\int_{-\infty}^{\infty} yf(x, y)\, dy}{f_X(x)}$$

quando $0 < f_X(x) < \infty$. Para outros valores definimos $E[\, Y \mid X = x\,] = 0$. A função definida por $m(x) = E[\, Y \mid X = x\,]$ chama-se em estatística *função de regressão* de Y sobre X.

Como veremos no Volume II, as expectâncias condicionais ocorrem em problemas estatísticos que envolvem predição e estimação bayesianas. De um ponto de vista mais geral, elas são também importantes na teoria avançada de probabilidade. Nos limitaremos a algumas aplicações elementares de expectâncias condicionais. A teoria geral é bastante sofisticada e não será requerida neste livro.

Exemplo 8. Suponha que X e Y têm a densidade conjunta do Exemplo 2 do Capítulo 6. Obtenha a expectância condicional de Y dado $X = x$.

No Exemplo 7 do Capítulo 6 vimos que a densidade condicional de Y dado $X = x$ é a densidade normal $n(x/2, 1)$ que sabemos ter média $x/2$. Assim

$$E[\, Y \mid X = x\,] = \frac{x}{2}.$$

Neste exemplo a variância condicional de Y dado $X = x$ é a constante 1.

Exemplo 9. Sejam X e Y variáveis aleatórias tendo densidade conjunta f dada por (5). Na seção precedente determinamos vários momentos envolvendo X e Y. Determinaremos aqui a densidade condicional e a expectância condicional de Y dado $X = x$.

A densidade marginal de X é dada por

$$f_X(x) = n(n-1) \int_x^1 (y - x)^{n-2}\, dy$$

$$= n(1 - x)^{n-1}, \quad 0 \leq x \leq 1,$$

e $f_X(x) = 0$ para outros valores de x. Assim para $0 \leqslant x < 1$,

$$f(y \mid x) = \begin{cases} \dfrac{(n-1)(y-x)^{n-2}}{(1-x)^{n-1}}, & x \leq y < 1, \\ 0, & \text{para outros valores de } y. \end{cases}$$

Conseqüentemente, para $0 \leqslant x < 1$,

$$\begin{aligned} E[Y \mid X = x] &= \int_{-\infty}^{\infty} y f(y \mid x)\, dy \\ &= (n-1)(1-x)^{1-n} \int_x^1 y(y-x)^{n-2}\, dy \\ &= (n-1)(1-x)^{1-n} \int_x^1 [(y-x)^{n-1} + x(y-x)^{n-2}]\, dy \\ &= (n-1)(1-x)^{1-n} \left[\frac{(1-x)^n}{n} + \frac{x(1-x)^{n-1}}{n-1}\right] \\ &= \frac{(n-1)(1-x)}{n} + x \\ &= \frac{n-1+x}{n}. \end{aligned}$$

Às vezes é conveniente calcular a expectância de Y de acordo com a fórmula

$$EY = \int_{-\infty}^{\infty} E[Y \mid X = x] f_X(x)\, dx. \tag{7}$$

Esta fórmula é uma decorrência imediata de (6). Para

$$\begin{aligned} \int_{-\infty}^{\infty} E[Y \mid X = x] f_X(x)\, dx &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} y f(x, y)\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y)\, dx\, dy \\ &= EY. \end{aligned}$$

Aplicando esta fórmula ao Exemplo 9, obtemos

$$\begin{aligned} EY &= \int_0^1 \left(\frac{n-1+x}{n}\right) n(1-x)^{n-1}\, dx \\ &= n \int_0^1 (1-x)^{n-1}\, dx - \int_0^1 (1-x)^n\, dx \\ &= 1 - \frac{1}{n+1} = \frac{n}{n+1}, \end{aligned}$$

que está de acordo com o resultado obtido no Exemplo 7.

Pode-se naturalmente definir de uma forma semelhante as expectâncias condicionais para variáveis aleatórias discretas.

7.5. O TEOREMA DO LIMITE CENTRAL

Através de toda esta seção $X_1, X_2, \ldots$ representarão variáveis aleatórias independentes, identicamente distribuídas com média μ e variância finita não-nula σ^2. Estaremos interessados em estudar a distribuição de $S_n = X_1 + \cdots + X_n$. Observe antes de mais nada que S_n tem média $n\mu$ e variância $n\sigma^2$.

Suponha a seguir que X_1 tem densidade f. Então S_n terá uma densidade f_{S_n} para todo $n \geqslant 1$, S_n. Como $f_{S_1} = f$, pode-se calcular sucessivamente as outras densidades usando as fórmulas obtidas nos Capítulos 3 e 6 para a densidade da soma de duas variáveis aleatórias independentes. Temos que

$$f_{S_n}(x) = \sum_y f_{S_{n-1}}(y) f(x - y)$$

ou

$$f_{S_n}(x) = \int_{-\infty}^{\infty} f_{S_{n-1}}(y) f(x - y)\, dy$$

conforme o tipo da variável aleatória X_1 discreta ou contínua. Para certas escolhas de f (por exemplo, binomial, binomial negativa, Poisson, normal e gama), podemos obter fórmulas simples para f_{S_n}. Entretanto, em geral temos que recorrer a métodos numéricos.

Um dos resultados mais importantes e extraordinários da teoria da probabilidade é que, para valores grandes de n, a distribuição de S_n depende essencialmente da distribuição de X_1 apenas através de μ e σ^2. Pode-se discutir mais facilmente este resultado em termos de variável aleatória normalizada

$$S_n^* = \frac{S_n - ES_n}{\sqrt{\operatorname{Var} S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

que tem média 0 e variância 1.

Para se ter uma idéia do comportamento da função de distribuição de S_n^* quando $n \to \infty$, consideremos inicialmente um caso especial em que a distribuição exata de S_n^* pode ser obtida facilmente. Suponha então que X_1 tem a distribuição normal de média μ e variância σ^2. Então, de acordo com os resultados do Capítulo 6, S_n^* distribui-se normalmente com média 0 e variância 1, ou em outras palavras, S_n^* tem a função de distribuição normal padrão Φ.

Suponha a seguir que X_1 assume os valores 1 e 0 com probabilidades p e $1 - p$, respectivamente. Então, como vimos no Capítulo 3, S_n tem a distribuição binomial de parâmetros n e p; isto é

$$P(S_n = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Foi descoberto por DeMoivre (1667-1754) e Laplace (1749-1827) que, neste caso, a função de distribuição de S_n^* se aproxima de Φ, a função de distribuição normal padrão, quando $n \to \infty$.

Houve mais recentemente diversas extensões do teorema limite de DeMoivre-Laplace, todas conhecidas como "teoremas do limite central". O mais simples e o mais conhecido desses resultados foi demonstrado por Lindeberg em 1922:

Teorema 3. Teorema do Limite Central. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas com média μ e variância finita não-nula σ^2. Seja $S_n = X_1 + \cdots + X_n$. Então

$$\lim_{n \to \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), \qquad -\infty < x < \infty. \tag{8}$$

A generalidade deste teorema é extraordinária. A variável aleatória X_1 pode ser discreta, contínua ou mista. Além do mais, a conclusão é válida mesmo que X_1 não tenha nenhum momento além do segundo. Outro aspecto bastante surpreendente do teorema é que a distribuição limite de S_n^* independe da distribuição específica de X_1 (naturalmente desde que as hipóteses do teorema sejam satisfeitas). Entretanto, não devemos ficar surpresos com o fato de que Φ é esta distribuição limite, pois vimos que isto é verdade quando X_1 tem distribuição normal ou binomial.

A demonstração do Teorema do Limite Central será adiada para o Capítulo 8, uma vez que ela requer técnicas avançadas ainda não discutidas que envolvem funções características. É possível dar uma demonstração elementar, porém, laboriosa do teorema de DeMoivre-Laplace, o caso especial do Teorema do Limite Central em que X_1 tem distribuição binomial. Existem maneiras elementares de tornar plausível o Teorema do Limite Central, mas elas não são demonstrações. Uma dessas maneiras é mostrar que se X_1 tem m-ésimo momento finito, então para qualquer número inteiro positivo m

$$\lim_{n \to \infty} E\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right)^m$$

existe e é igual ao m-ésimo momento da distribuição normal padrão.

Neste estágio é mais interessante compreender o significado do Teorema do Limite Central e ver como podemos aplicá-lo em situações típicas.

Exemplo 10. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas segundo uma distribuição de Poisson de parâmetro λ. Então, de acordo com os resultados do Capítulo 4, $\mu = \sigma^2 = \lambda$ e S_n tem uma distribuição de Poisson de parâmetro $n\lambda$. O Teorema do Limite Central implica em

$$\lim_{n \to \infty} P\left(\frac{S_n - n\lambda}{\sqrt{n\lambda}} \leq x\right) = \Phi(x), \qquad -\infty < x < \infty.$$

Pode-se estender o resultado deste exemplo e mostrar que se X_t é uma variável aleatória com distribuição de Poisson de parâmetro $\lambda = t$, então

$$(9) \qquad \lim_{t \to \infty} P\left(\frac{X_t - EX_t}{\sqrt{\operatorname{Var} X_t}} \le x\right) = \Phi(x), \qquad -\infty < x < \infty.$$

A Equação (9) é também válida para o caso em que X_t tem uma distribuição gama $\Gamma(t, \lambda)$ com λ fixo ou uma distribuição binomial negativa de parâmetros $\alpha = t$ e p fixo.

7.5.1. APROXIMAÇÕES NORMAIS.

O Teorema do Limite Central sugere fortemente que para n grande devemos fazer a aproximação

$$P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) \approx \Phi(x), \qquad -\infty < x < \infty,$$

ou equivalentemente

$$(10) \qquad P(S_n \le x) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$$

$$= \Phi\left(\frac{x - ES_n}{\sqrt{\operatorname{Var} S_n}}\right), \qquad -\infty < x < \infty.$$

Nós nos referimos a (10) como uma fórmula de *aproximação normal*. De acordo com esta fórmula, aproximamos a função de distribuição de S_n pela função de distribuição normal de mesma média e variância. Uma dificuldade em aplicar a fórmula de aproximação normal é decidir sobre a ordem de grandeza de n para que (10) seja válida com um desejado grau de precisão. Diversos estudos numéricos indicam que em aplicações práticas típicas, $n = 25$ é suficientemente grande para que (10) seja válida.

Como um exemplo em que a aproximação normal é aplicável, sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, cada uma com densidade exponencial de parâmetro $\lambda = 1$. Então (10) transforma-se em

$$(11) \qquad P(S_n \le x) \approx \Phi\left(\frac{x - n}{\sqrt{n}}\right), \qquad -\infty < x < \infty.$$

Gráficos mostrando a precisão desta aproximação são dados na Figura 1 para $n = 10$.

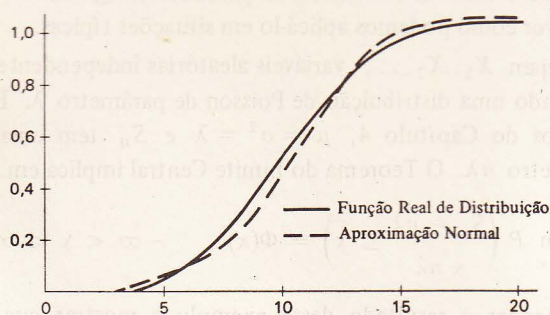



Figura 1

Exemplo 11. Suponha que a duração de um certo tipo de lâmpada depois de instalada distribui-se, exponencialmente, com duração média de 10 dias. Quando uma lâmpada queima, instala-se outra do mesmo tipo em seu lugar. Obtenha a probabilidade de que sejam necessárias mais de 50 lâmpadas durante o período de um ano.

Para resolver este problema representamos por X_n a duração da n-ésima lâmpada que é instalada. Supomos que $X_1, X_2, \ldots$ são variáveis aleatórias independentes com densidade exponencial de média 10 ou parâmetro $\lambda = 1/10$. Então $S_n = X_1 + \cdots + X_n$ representa o tempo em que se queima a n-ésima lâmpada. Desejamos determinar $P(S_{50} < 365)$. Sabemos que S_{50} tem média $50\lambda^{-1} = 500$ e variância $50\lambda^{-2} = 5000$. Assim, pela fórmula de aproximação normal (10)

$$P(S_{50} < 365) \approx \Phi\left(\frac{365 - 500}{\sqrt{5000}}\right)$$

$$= \Phi(-1{,}91) = 0{,}028.$$

É portanto bastante improvável que sejam necessárias mais de 50 lâmpadas.

Suponha que S_n é uma variável aleatória contínua com densidade f_{S_n}. Diferenciando os têrmos de (10) obtemos

$$(12) \qquad f_{S_n}(x) \approx \frac{1}{\sigma\sqrt{n}}\,\varphi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right), \qquad -\infty < x < \infty.$$

Embora, a derivação de (12) esteja longe de ser uma demonstração, (12) é uma aproximação boa para n grande (sob a condição branda adicional de que f_{S_n} seja uma função limitada para algum n).

Como um exemplo desta aproximação, suponha que X_1 se distribui exponencialmente com parâmetros $\lambda = 1$, de modo que (11) é aplicável. Então (12) transforma-se em

$$(13) \qquad f_{S_n}(x) \approx \frac{1}{\sqrt{n}}\,\varphi\left(\frac{x - n}{\sqrt{n}}\right), \qquad -\infty < n < \infty.$$

Gráficos mostrando a precisão desta aproximação são dados na Figura 2 para $n = 10$.

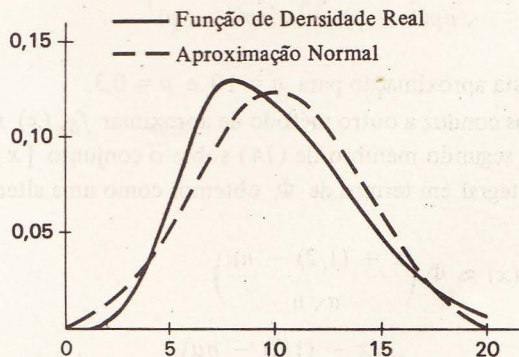



Figura 2

As formas do Teorema do Limite Central que envolvem densidades em lugar das funções de distribuição são conhecidas como teoremas do limite central "local". Elas são também importantes, especialmente na teoria avançada da probabilidade.

Existe uma aproximação semelhante a (12) para variáveis aleatórias discretas. Naturalmente o estabelecimento preciso de tal aproximação depende da natureza dos valores possíveis de S_n, isto é, aqueles valores de x tal que $f_{S_n}(x) = P(S_n = x) > 0$. Por simplicidade fazemos as duas hipóteses seguintes:

(i) Se x é um valor possível de X_1, então x é inteiro;

(ii) Se a é um valor possível de X_1, então o máximo divisor comum do conjunto

$$\{ x - a \mid x \text{ é um valor possível de } X_1 \}$$

é um.

Excluímos, por exemplo, a variável aleatória X_1 tal que $P(X_1 = 1) = P(X_1 = 3) = 1/2$, pois neste caso o máximo divisor comum do conjunto indicado é 2. Sob as hipóteses (i) e (ii), a aproximação

$$(14) \qquad f_{S_n}(x) \approx \frac{1}{\sigma\sqrt{n}}\,\varphi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right), \qquad x \text{ inteiro}$$

é válida para n grande.

Exemplo 12. Seja X_1 uma variável aleatória que assume os valores 1 e 0 com probabilidades p e $1 - p$, respectivamente. Então (i) e (ii) são satisfeitas e (14) é aplicável com $\mu = p$ e $\sigma^2 = p(1 - p)$. Como S_n tem a distribuição binomial de parâmetros n e p, temos a aproximação

$$(15) \qquad f_{S_n}(x) = \binom{n}{x} p^x(1 - p)^{n-x}$$

$$\approx \frac{1}{\sqrt{np(1 - p)}}\,\varphi\left(\frac{x - np}{\sqrt{np(1 - p)}}\right), \qquad x \text{ inteiro}$$

A Figura 3 mostra esta aproximação para $n = 10$ e $p = 0{,}3$.

A Figura 3 nos conduz a outro método de aproximar $f_{S_n}(x)$ no caso discreto, isto é, a integral do segundo membro de (14) sobre o conjunto $[x - 1/2, x + 1/2]$. Expressando essa integral em termos de Φ, obtemos como uma alternativa para (14)

$$(16) \qquad f_{S_n}(x) \approx \Phi\left(\frac{x + (1/2) - n\mu}{\sigma\sqrt{n}}\right)$$

$$- \Phi\left(\frac{x - (1/2) - n\mu}{\sigma\sqrt{n}}\right), \qquad x \text{ inteiro.}$$

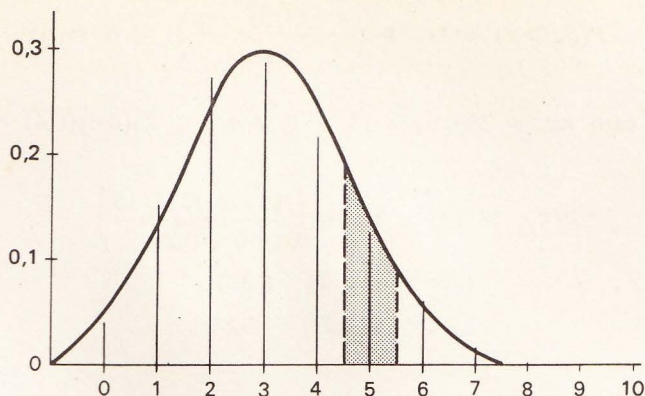



Figura 3

A área da região sombreada da Figura 3 é uma aproximação para $P(S_n = 5)$.

Finalmente somando (16) sobre o conjunto $\{\ldots, x-2, x-1, x\}$

$$(17) \qquad P(S_n \leqslant x) \approx \Phi\left(\frac{x + (1/2) - n\mu}{\sigma\sqrt{n}}\right), \quad x \text{ inteiro},$$

Quando S_n é discreto e as condições (i) e (ii) se verificam, (17) é em geral mais precisa do que a fórmula original de aproximação normal (10). Na Figura 4 comparamos as aproximações das fórmulas (10) e (17) quando S_n tem a distribuição binomial de parâmetros $n = 10$ e $p = 0{,}3$.

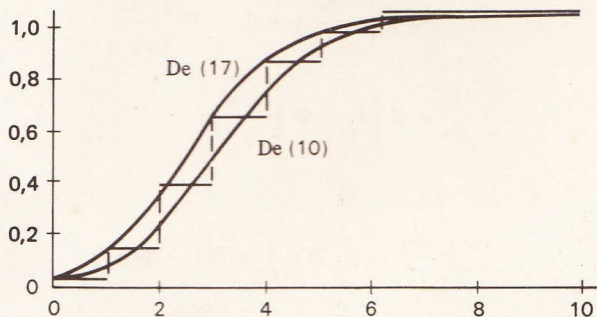



Figura 4

Exemplo 13. Um certo jogador de basquetebol sabe que converterá em média 60% dos seus lances-livres. Qual a probabilidade de que ele tenha êxito em mais da metade das vezes em 25 lances-livres?

Suponhamos que a implicação do problema é que o número S_n de sucessos em n lances-livres distribui-se binomialmente com parâmetros n e $p = 0{,}6$. Como $P(S_n \geqslant x) = 1 - P(S_n \leqslant x - 1)$ obtemos de (17) a aproximação

$$(18) \qquad P(S_n \geqslant x) \approx 1 - \Phi\left(\frac{x - (1/2) - n\mu}{\sigma\sqrt{n}}\right), \qquad x \text{ inteiro.}$$

Para o nosso caso $n\mu = 25(0{,}6) = 15$ e $\sigma\sqrt{n} = \sqrt{25(0{,}6)(0{,}4)} = 5\sqrt{0{,}24}$.
Assim

$$\begin{aligned} P(S_{25} \geqslant 13) &\approx 1 - \Phi\left(\frac{13 - (1/2) - 15}{5\sqrt{0{,}24}}\right) \\ &= 1 - \Phi(-1{,}02) \\ &= \Phi(1{,}02) = 0{,}846. \end{aligned}$$

7.5.2. APLICAÇÃO À AMOSTRAGEM. Pode-se considerar o Teorema do Limite Central e as correspondentes fórmulas de aproximação normal como refinamentos da Lei Fraca dos Grandes Números discutida no Capítulo 4. Lembramos que esta lei estabelece que para n grande S_n/n deve estar próximo de μ com probabilidade 1. Entretanto, a lei fraca por si só não fornece informação alguma quanto a precisão de tal estimativa. Como vimos no Capítulo 4, a Densidade de Chebyshev lança alguma luz sobre esta questão.

A fórmula de aproximação normal (10) é também útil neste contexto. Para $c > 0$

$$\begin{aligned} P\left(\left|\frac{S_n}{n} - \mu\right| \geq c\right) &= P(S_n \leq n\mu - nc) + P(S_n \geq n\mu + nc) \\ &\approx \Phi\left(\frac{-nc}{\sigma\sqrt{n}}\right) + 1 - \Phi\left(\frac{nc}{\sigma\sqrt{n}}\right) \\ &= 2\left[1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right]. \end{aligned}$$

Em outras palavras

$$(19) \qquad P\left(\frac{S_n}{n} - \mu \geqslant c\right) \approx 2(1 - \Phi(\delta)),$$

onde

$$(20) \qquad \delta = \frac{c\sqrt{n}}{\sigma}.$$

Exemplo 14. Decide-se tomar uma amostra de tamanho n para determinar a porcentagem de eleitores que tencionam votar em determinado candidato em uma eleição. Seja $X_k = 1$ se o k-ésimo elemento da amostra tenciona votar no candidato e $X_k = 0$ caso contrário. Supomos que $X_1, \ldots, X_n$ sejam variáveis aleatórias independentes, identicamente distribuídas, de tal sorte que $P(X_1 = 1) = p$ e

$P(X_1 = 0) = 1 - p$. Então $\mu = p$ e $\sigma^2 = p(1-p)$. Supomos também que o valor de p seja suficientemente próximo de 0,5 para que $\sigma = \sqrt{p(1-p)}$ possa ser aproximado satisfatoriamente por $\sigma \approx 1/2$ (observe que σ tem um máximo de $1/2$ para $p = 0,5$ e que σ permanece acima de $0,3 \leqslant p \leqslant 0,7$, quando σ varia de 0,458, o que é suficientemente próximo de 1/2). A variável aleatória S_n/n representa a fração de pessoas na amostra que tencionam votar no candidato, e pode ser usada para estimar a proporção verdadeira porém desconhecida, p. Usaremos as aproximações normais para resolver os três problemas seguintes:

(i) Suponha que $n = 900$. Obtenha a probabilidade de

$$\left|\frac{S_n}{n} - p\right| \geqslant 0,025.$$

(ii) Suponha que $n = 900$. Determine c tal que

$$P\left(\left|\frac{S_n}{n} - p\right| \geqslant c\right) = 0,01.$$

(iii) Determine n tal que

$$P\left(\left|\frac{S_n}{n} - p\right| \geqslant 0,025\right) = 0,01.$$

Solução de (i). De acordo com (20)

$$\delta = \frac{(0,025)\sqrt{900}}{0,5} = 1,5,$$

de modo que em virtude de (19)

$$P\left(\left|\frac{S_n}{n} - p\right| \geqslant 0,025\right) \approx 2(1 - \Phi(1,5))$$

$$= 2(0,067) = 0,134.$$

Solução de (ii). Escolhemos inicialmente δ de tal forma que $2(1 - \Phi(\delta)) = 0,1$ ou $\Phi(\delta) = 0,995$. Uma inspeção da Tabela I mostra que $\delta = 2,58$. Resolvendo (20) em c obtemos

$$c = \frac{\delta\sigma}{\sqrt{n}} = \frac{(2,58)(0,5)}{\sqrt{900}} = 0,043.$$

Solução de (iii). Com em (ii), $\delta = 2,58$. Resolvendo (20) em n, obtemos

$$n = \frac{\delta^2 \sigma^2}{c^2} = \frac{(2{,}58)^2\,(0{,}25)}{(0{,}025)^2} = 2663.$$

Vale a pena comparar os resultados de (ii) e (iii). Em ambos os casos

$$P\left(\left|\frac{S_n}{n} - p\right| \geq c\right) \approx 0{,}01.$$

Temos $c = 0{,}043$ e $n = 900$ em (ii), enquanto $c = 0{,}025$ e $n = 2663$ em (iii). Assim, para reduzir c proporcionalmente ao fator $43/25$, somos forçados a aumentar n proporcionalmente ao quadrado deste fator. Isto é válido em geral sempre que desejamos manter

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq c\right)$$

constante. Pois então δ é determinado por (19) e, de acordo com (20), n está relacionado com c pela expressão $n = \delta^2 \sigma^2 / c^2$. No mesmo contexto, se aumentamos n proporcionalmente a um dado fator, reduzimos c proporcionalmente à raiz quadrada do mesmo fator.

Exercícios

1. Suponha que X tem uma densidade Beta de parâmetros α_1 e α_2. Obtenha EX.
2. Sejam X e Y aleatórias independentes tendo as densidades gama $\Gamma(\alpha_1, \lambda)$ e $\Gamma(\alpha_2, \lambda)$ respectivamente. Seja $Z = Y/X$. Para que valores de α_1 e α_2 terá Z uma expectância finita? Obtenha EZ quando ela existe. *Sugestão:* veja o Teorema 3 do Capítulo 6 e discussão sobre o mesmo.
3. Suponha que X tem a densidade normal $n(0, \sigma^2)$. Obtenha $E\,|X|$. *Sugestão:* use o resultado do Exercício 31 do Capítulo 5.
4. Suponha que X tem uma densidade exponencial de parâmetro λ e que X_ϵ seja definido em termos de X e $\epsilon > 0$ através de (1). Qual é a distribuição de X_ϵ/ϵ? Obtenha EX_ϵ e determine o seu limite quando $\epsilon \to 0$.
5. Suponha que X tem uma densidade Beta de parâmetros α_1 e α_2. Obtenha os momentos e a variância de X.
6. Suponha que X tem uma distribuição χ^2 com n graus de liberdade. Obtenha a média de $Y = \sqrt{X}$.
7. Seja X a variável aleatória do Exemplo 7. Obtenha EX^m a partir da densidade marginal f_X.
8. Seja Z a variável aleatória do Exercício 2. Obtenha a variância de Z.

9. Sejam U_1 e U_2 variáveis aleatórias independentes que têm densidade exponencial de parâmetro λ e seja $Y = \max(U_1, U_2)$. Obtenha a média e variância de Y (ver Seção 6.5).

10. Seja X a variável aleatória do Exemplo 1 do Capítulo 5. Obtenha a média e variância de X.

11. Seja X a variável aleatória do Exemplo 1 do Capítulo 6. Obtenha a média e variância de X. *Sugestão:* reduza EX^2 a uma integral Beta.

12. Obtenha a média e variância da variável aleatória Z do Exercício 17 do Capítulo 6.

13. Obtenha a média e variância da variável aleatória Y do Exercício 28 do Capítulo 6.

14. Seja X o seno de um ângulo em radianos escolhido uniformemente em $(-\pi/2, \pi/2)$. Obtenha a média e variância de X.

15. Suponha que X tem a densidade normal $n(0, \sigma^2)$. Obtenha a média e variância das seguintes variáveis aleatórias:

    (a) $|X|$;

    (b) X^2;

    (c) e^{tX}.

16. Suponha que X tem a densidade gama $\Gamma(\alpha, \lambda)$. Determine os valores de t para os quais e^{tX} tem expectância finita. Obtenha Ee^{tX} para estes valores de t.

17. Suponha que X tem a densidade $\Gamma(\alpha, \lambda)$. Determine os valores de r para os quais X^r tem expectância finita. Obtenha EX^r para estes valores de r.

18. Seja X uma variável aleatória contínua não-negativa que tem densidade f e função de distribuição F. Mostre que X tem expectância finita se e somente se

$$\int_0^\infty (1 - F(x))\, dx < \infty$$

e então

$$EX = \int_0^\infty (1 - F(x))\, dx.$$

Sugestão: veja a demonstração do Teorema 5 do Capítulo 4.

19. Seja X_k o k-ésimo estatístico de ordem de uma amostra das variáveis aleatórias $U_1, \ldots, U_n$ que são independentes e se distribuem uniformemente em $(0, 1)$. Obtenha a média e variância da X_k.

20. Sejam X e Y as variáveis aleatórias do Exemplo 7 e seja $R = Y - X$. Obtenha a média e variância de R. *Sugestão:* use a Equação (16) do Capítulo 4.

21. Suponha que X e Y tem a densidade f dada no Exercício 9 do Capítulo 6. Obtenha a correlação entre X e Y.

22. Sejam X e Y variáveis independentes tais que X tem a densidade normal $n(\mu, \sigma^2)$ e Y a densidade gama $\Gamma(\alpha, \lambda)$. Obtenha a média e variância de $Z = XY$.

23. Sejam X e Y variáveis aleatórias que têm média 0, variância 1 e correlação ρ. Mostre que $X - \rho Y$ e Y não são correlacionadas e que $X - \rho Y$ tem média 0 e variância $1 - \rho^2$.

24. Sejam X, Y e Z variáveis aleatórias que têm média 0 e variância 1. Seja ρ_1 a correlação entre X e Y, ρ_2 a correlação entre Y e Z e ρ_3 a correlação entre X e Z. Mostre que

$$\rho_2 \geqslant \rho_1 \rho_2 - \sqrt{1 - \rho_1^2}\,\sqrt{1 - \rho_2^2}.$$

Sugestão: escreva

$$XZ = [\rho_1 Y + (X - \rho_1 Y)][\rho_2 Y + (Z - \rho_2 Y)],$$

e use o Exercício 23 e a desigualdade de Schwarz.

25. Sejam X, Y e Z como no exercício anterior. Suponha que $\rho_1 \geqslant 0{,}9$ e $\rho_2 \geqslant 0{,}8$. O que se pode dizer acerca de ρ_3?

26. Suponha que X e Y têm uma densidade f uniforme no interior do triângulo com vértices em $(0, 0)$, $(2, 0)$ e $(1, 2)$. Obtenha a expectância condicional de Y dado X.

27. Sejam X e Y variáveis aleatórias independentes tendo as densidades gama $\Gamma(\alpha_1, \lambda)$ e $\Gamma(\alpha_2, \lambda)$, respectivamente, e seja $Z = X + Y$. Obtenha a expectância condicional de X dado Z.

28. Sejam Π e Y as variáveis aleatórias do Exercício 26 do Capítulo 6. Obtenha a expectância condicional de Π dado Y.

29. Sejam X e Y variáveis aleatórias contínuas tendo uma densidade conjunta. Suponha que Y e $\varphi(X)Y$ têm expectâncias finitas. Mostre que

$$E\varphi(X)Y = \int_{-\infty}^{\infty} \varphi(x) E[Y \mid X = x] f_X(x)\, dx.$$

30. Sejam X e Y variáveis aleatórias contínuas tendo uma densidade conjunta, e seja $\operatorname{Var}[Y \mid X = x]$ a variância condicional de Y dado $X = x$. Mostre que se $E[Y \mid X = x] = \mu$ independentemente de X. Então $EY = \mu$ e

$$\operatorname{Var} Y = \int_{-\infty}^{\infty} \operatorname{Var}[Y \mid X = x] f_X(x)\, dx.$$

31. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas com média 0 e variância finita não nula σ^2 e seja $S_n = X_1 + \cdots + X_n$. Mostre que se X_1 tem terceiro momento finito, então $ES_n^3 = nEX_1^3$ e

$$\lim_{n\to\infty} E\left(\frac{S_n}{\sigma\sqrt{n}}\right)^3 = 0,$$

que é o terceiro momento da distribuição normal padrão.

32. Sejam $X_1, \ldots, X_n$ e S_n como no Exercício 31. Mostre que se X_1 tem quarto momento finito, então

$$ES_n^4 = nEX_1^4 + 3n(n-1)\sigma^4$$

e

$$\lim_{n\to\infty} E\left(\frac{S_n}{\sigma\sqrt{n}}\right)^4 = 3,$$

que é o quarto momento da distribuição normal padrão. *Sugestão:* o termo $3n(n-1)$ vem da expressão

$$\binom{n}{2}\frac{4!}{2!\,2!}.$$

33. Suponha que X tem a densidade gama $\Gamma(\alpha, \lambda)$. Obtenha a aproximação normal para $P(X \leqslant x)$.

34. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com distribuição normal de média 0 e variância σ^2.

    (a) Quais são a média e variância da variável aleatória X_1^2?

    (b) Como se deve aproximar $P(X_1^2 + \cdots + X_n^2 \leqslant x)$ em termos de Φ?

35. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes com distribuição normal de média 0 e variância 1 (ver exercício anterior).

    (a) Determine $P(X_1^2 + \cdots + X_{100}^2 \leqslant 120)$.

    (b) Determine $P(80 \leqslant X_1^2 + \cdots + X_{100}^2 \leqslant 120)$.

    (c) Determine c tal que $P(X_1^2 + \cdots + X_{100}^2 \leqslant 100 + c) = 0{,}95$.

    (d) Determine c tal que $P(100 - c \leqslant X_1^2 + \cdots + X_{100}^2 \leqslant 100 + c) = 0{,}95$.

36. Um corredor procura controlar seus passos em uma corrida de 100 metros. Seus passos distribuem-se independentemente com média $\mu = 0{,}97$ metros e desvio padrão $\sigma = 0{,}1$ metros. Determine a probabilidade de que 100 passos difiram de 100 metros por não mais de 5 metros.

37. Arredonda-se vinte números para o inteiro mais próximo e soma-se os números resultantes. Suponha que os erros individuais de arredondamento são independentes e se distribuem uniformemente em $(-1/2, 1/2)$. Determine a probabilidade de que a soma obtida difira da soma dos vinte números originais por mais de 3.

38. Lança-se uma moeda equilibrada até observar 100 caras. Determine a probabilidade de que sejam necessários 226 lançamentos no mínimo.

39. No exercício anterior determine a probabilidade de que sejam necessários exatamente 226 lançamentos.

40. Suponha que X tem uma distribuição de Poisson de parâmetro λ.

    (a) Como se deve aproximar $f_X(x)$ em termos da densidade normal padrão φ?

    (b) Como se deve aproximar $f_X(x)$ em termos da função de distribuição normal padrão Φ?

41. Suponha que S_n tem uma distribuição binomial de parâmetros n e $p = 1/2$. Como se comporta $P(S_{2n} = n)$ para n grande? *Sugestão:* use a aproximação (15).

42. Os jogadores A e B fazem uma série de apostas de \$ 1 em que cada jogador tem probabilidade 1/2 de ganhar. Seja S_n a quantia ganha pelo jogador A após n apostas. Como se comporta $P(S_{2n} = 0)$ para n grande? *Sugestão:* veja o problema anterior. Por que (15) não é diretamente aplicável neste caso?

43. Os candidatos A e B concorrem a um cargo e, 55% do eleitorado apoia o candidato B. Qual é a probabilidade de que em uma amostra de tamanho 100, pelo menos a metade dos entrevistados apoiem o candidato A ?

44. Uma firma de pesquisa de opinião pública entrevista 1200 eleitores para estimar a proporção dos que planejam votar no candidato A. Qual deve ser a verdadeira proporção p para que o candidato A possa estar 95% certo de que a maioria dos entrevistados votarão nele?

45. Suponha que o candidato A do problema anterior insiste que o tamanho da amostra deve ser tal que se 51% de todos os eleitores o favorecem, ele pode estar 95% certo de obter a maioria dos votos na amostra. Qual deve ser o valor de n?

46. Resolva o Exercício 27 do Capítulo 4 usando a aproximação normal.

8

FUNÇÕES GERATRIZES DE MOMENTOS E FUNÇÕES CARACTERÍSTICAS

Alguns dos instrumentos mais importantes na teoria da probabilidade provêm de outros ramos da matemática. Discutiremos neste capítulo dois de tais instrumentos intimamente relacionados. Começaremos com funções geratrizes de momentos e a seguir trataremos de funções características. Estas são de compreensão um pouco mais difícil em nível elementar porque requerem o uso de números complexos. Entretanto vale a pena superar este obstáculo, pois o conhecimento das propriedades das funções características nos permitirá demonstrar tanto a Lei Fraca dos Grandes Números como o Teorema do Limite Central (Seção 8.4).

8.1. FUNÇÕES GERATRIZES DE MOMENTOS

Define-se a *função geratriz de momentos* $M_X(t)$ de uma variável aleatória X através de

$$M_X(t) = Ee^{tX}.$$

O domínio de M_X é constituído por todo real t para o qual e^{tX} tem expectância finita.

Exemplo 1. Seja X uma variável aleatória com distribuição normal de média μ e variância σ^2. Então

$$\begin{aligned} M_X(t) = Ee^{tX} &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} dx \\ &= \int_{-\infty}^{\infty} e^{t(y+\mu)} \frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2} dy \\ &= e^{\mu t} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{ty-(y^2/2\sigma^2)} dy. \end{aligned}$$

Mas

$$ty - \frac{y^2}{2\sigma^2} = -\frac{(y-\sigma^2 t)^2}{2\sigma^2} + \frac{\sigma^2 t^2}{2}.$$

Conseqüentemente

$$M_X(t) = e^{\mu t} e^{\sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-[(y-\sigma^2 t)^2/2\sigma^2]} dt.$$

Como a última integral representa a integral da densidade normal $n(\sigma^2 t, \sigma^2)$, seu valor é um e portanto

(1) $$M_X(t) = e^{\mu t} e^{\sigma^2 t^2/2}, \quad -\infty < t < \infty.$$

Exemplo 2. Suponha que X tem a densidade gama de parâmetros α e λ. Então

$$M_X(t) = \int_0^{\infty} e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-(\lambda-t)x} dx$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(\lambda - t)^\alpha}$$

para $-\infty < t < \lambda$. A integral diverge para $\lambda \leqslant t < \infty$. Assim

(2) $$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha, \quad -\infty < t < \lambda.$$

Suponha agora que X é uma variável aleatória discreta cujos valores possíveis são todos inteiros não-negativos. Então

$$M_X(t) = \sum_{n=0}^{\infty} e^{nt} P(X = n).$$

No Capítulo 3 definimos a função geratriz de probabilidades de tais variáveis aleatórias como

$$\Phi_X(t) = \sum_{n=0}^{\infty} t^n P(X = n).$$

Das duas fórmulas acima torna-se claro que

(3) $$M_X(t) = \Phi_X(e^t).$$

A fórmula (3) nos permite obter a função geratriz de momento diretamente da função geratriz de probabilidade. Por exemplo, se X tem uma distribuição binomial de parâmetros n e p, então como foi visto no Exemplo 16 do Capítulo 3,

$$\Phi_X(t) = (pt + 1 - p)^n.$$

Segue-se imediatamente que

$$M_X(t) = (pe^t + 1 - p)^n.$$

Analogamente, se X tem uma distribuição de Poisson de parâmetro λ, então de acordo com o Exemplo 18 do Capítulo 3,

$$\Phi_X(t) = e^{\lambda(t-1)}.$$

Conseqüentemente

$$M_X(t) = e^{\lambda(e^t - 1)}.$$

Naturalmente nestes dois exemplos $M_X(t)$ poderia também ser facilmente obtido diretamente da definição de função geratriz de momento.

Se X e Y são variáveis aleatórias independentes, e^{tX} e e^{tY} são também independentes. Conseqüentemente

$$M_{X+Y}(t) = Ee^{t(X+Y)} = Ee^{tX}e^{tY} = Ee^{tX}Ee^{tY}$$
$$= M_X(t)M_Y(t).$$

Segue-se facilmente que se $X_1, \ldots, X_n$ são independentes e têm a mesma distribuição, então

(4) $$M_{X_1 + \cdots + X_n}(t) = (M_{X_1}(t))^n.$$

Para ver a razão pela qual chamamos $M_X(t)$ função geratriz de momento, escrevemos

$$M_X(t) = Ee^{tX} = E \sum_{n=0}^{\infty} \frac{t^n X^n}{n!}.$$

Suponha que $M_X(t)$ é finito em $-t_0 < t < t_0$ para algum número positivo t_0. Neste caso pode-se mostrar que é lícito inverter a ordem de expectância e somatório na última expressão de $M_X(t)$. Em outras palavras

(5) $$M_X(t) = \sum_{n=0}^{\infty} \frac{EX^n}{n!} t^n$$

para $-t_0 < t < t_0$. Em particular, (5) é válido para todo t se $M_X(t)$ é finito para todo t. A série de Taylor por $M_X(t)$ é

(6) $$M_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \frac{d^n}{dt^n} M_X(t) \Big|_{t=0}.$$

Comparando os coeficientes de t^n em (5) e (6), vemos que

(7) $$EX^n = \frac{d^n}{dt^n} M_X(t) \Big|_{t=0}.$$

Exemplo 3. Suponha que X se distribui normalmente com média 0 e variância σ^2. Use a função geratriz de momentos para obter os momentos de X.

Observe inicialmente de (1) que

$$M_X(t) = e^{\sigma^2 t^2/2} = \sum_{n=0}^{\infty} \left(\frac{\sigma^2 t^2}{2}\right)^n \frac{1}{n!}$$

$$= \sum_{n=0}^{\infty} \frac{\sigma^{2n}}{2^n n!} t^{2n}.$$

Assim os momentos de ordem ímpar de X são iguais a zero, e os momentos de ordem par são dados por

$$\frac{EX^{2n}}{(2n)!} = \frac{\sigma^{2n}}{2^n n!}$$

ou

$$EX^{2n} = \frac{\sigma^{2n}(2n)!}{2^n n!}.$$

Isto está de acordo com os resultados obtidos no Capítulo 7.

Este exemplo pode ser usado para ilustrar (7). Uma vez que

$$\frac{d}{dt} e^{\sigma^2 t^2/2} = \sigma^2 t e^{\sigma^2 t^2/2}$$

e

$$\frac{d^2}{dt^2} e^{\sigma^2 t^2/2} = (\sigma^2 + \sigma^4 t^2) e^{\sigma^2 t^2/2},$$

segue-se que

$$\left.\frac{d}{dt} e^{\sigma^2 t^2/2}\right|_{t=0} = 0$$

e

$$\left.\frac{d^2}{dt^2} e^{\sigma^2 t^2/2}\right|_{t=0} = \sigma^2,$$

que são simplesmente os dois primeiros momentos de X.

8.2. FUNÇÕES CARACTERÍSTICAS

Define-se a *função característica* de uma variável aleatória X como

$$\varphi_X(t) = Ee^{itX}, \qquad -\infty < t < \infty,$$

onde $i = \sqrt{-1}$. As funções características são um pouco mais complicadas do que as funções geratrizes de momento, na medida em que envolvem números complexos. Entretanto elas possuem duas vantagens importantes em relação as funções

geratrizes de momentos. Em primeiro lugar $\varphi_X(t)$ é finito para todas as variáveis aleatórias X e todos os números reais t. Em segundo lugar, a função de distribuição de X e em geral a função de densidade, quando existe, podem ser obtidas da função característica através de uma "fórmula de inversão". Usando as propriedades das funções características poderemos demonstrar a Lei Fraca dos Grandes Números e o Teorema do Limite Central, o que não era possível através de funções geratrizes de momento.

Antes de discutir as funções características faremos um breve resumo de alguns fatos complexos, envolvendo variáveis, que serão necessários.

Podemos escrever qualquer número complexo z na forma $z = x + iy$, onde x e y são números reais. Define-se o valor absoluto $|z|$ de um número complexo z através de $|z| = (x^2 + y^2)^{1/2}$. Define-se a distância entre dois números complexos z_1 e z_2 como sendo $|z_1 - z_2|$.

Se a função de uma variável real tem uma expansão em série de potência com um raio de convergência positivo, podemos usar esta série de potência para definir uma função correspondente de uma variável complexa. Assim definimos

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

para qualquer número complexo z. A relação

$$e^{z_1 + z_2} = e^{z_1} e^{z_2}$$

permanece válida para todos os números complexos z_1 e z_2. Fazendo $z = it$, onde t é um número real, vemos que

$$\begin{aligned} e^{it} &= \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \\ &= \left(1 + it - \frac{t^2}{2} - \frac{it^3}{3!} + \frac{t^4}{4!} + \frac{it^5}{5!} - \cdots\right) \\ &= \left(1 - \frac{t^2}{2!} + \frac{t^4}{4!} - \cdots\right) + i\left(t - \frac{t^3}{3!} + \frac{t^5}{5!} - \cdots\right). \end{aligned}$$

Como as duas séries nesta última expressão correspondem a $\cos t$ e $\operatorname{sen} t$, segue-se que

$$e^{it} = \cos t + i \operatorname{sen} t. \tag{8}$$

Usando a fato de que $\cos(-t) = \cos t$ e $\operatorname{sen}(-t) = -\operatorname{sen} t$, vemos que

$$e^{-it} = \cos t - i \operatorname{sen} t.$$

Dessas fórmulas podemos obter $\cos t$ e $\operatorname{sen} t$:

(9)
$$\cos t = \frac{e^{it} + e^{-it}}{2} \quad \text{e} \quad \text{sen}\, t = \frac{e^{it} - e^{-it}}{2i}$$

Segue-se também de (8) que

$$|e^{it}| = (\cos^2 t + \text{sen}^2\, t)^{1/2} = 1.$$

Se $f(t)$ e $g(t)$ são funções reais de t, então $h(t) = f(t) + ig(t)$ define uma função complexa de t. Podemos diferenciar $h(t)$ diferenciando $f(t)$ e $g(t)$ separadamente, isto é,

$$h'(t) = f'(t) + ig'(t),$$

desde que $f'(t)$ e $g'(t)$ existem. Da mesma forma definimos

$$\int_a^b h(t)\, dt = \int_a^b f(t)\, dt + i \int_a^b g(t)\, dt,$$

desde que as integrais indicadas envolvendo f e g existem. A fórmula

$$\frac{d}{dt} e^{ct} = ce^{ct}$$

é válida para qualquer constante complexa c. O teorema fundamental de cálculo continua válido e, em particular, se c é uma constante complexa não-nula, então

$$\int_a^b e^{ct}\, dt = \frac{e^{cb} - e^{ca}}{c}$$

Pode-se escrever uma variável aleatória complexa Z na forma $Z = X + iY$, onde X e Y são variáveis aleatórias reais. Define-se sua expectância EZ como

$$EZ = E(X + iY) = EX + iEY$$

sempre que EX e EY forem bem definidas. Como no caso das variáveis aleatórias reais, Z tem expectância finita se, e somente se, $E|Z| < \infty$, e neste caso

$$|EZ| \leqslant E|Z|$$

A fórmula

$$E(a_1 Z_1 + a_2 Z_2) = a_1 EZ_1 + a_2 EZ_2$$

é válida quando a_1 e a_2 forem constantes complexas e Z_1 e Z_2 forem variáveis aleatórias complexas com expectância finita.

Representamos por X e Y, com ou sem subscritos, as variáveis aleatórias reais. Assim, na frase "seja X uma variável aleatória..." fica subentendido que se trata de uma variável aleatória real.

Suponha então que X é uma variável aleatória e t é uma constante (reservamos o símbolo t para constantes reais). Então, $|e^{itX}| = 1$, de modo que e^{itX} tem expectância finita e a função característica $\varphi_X(t)$, $-\infty < t < \infty$, dada por

$$\varphi_X(t) = Ee^{itX}, \qquad -\infty < t < \infty,$$

é bem definida. Vemos que $\varphi_X(0) = Ee^0 = E1 = 1$ e, para $-\infty < t < \infty$,

$$|\varphi_X(t)| = |Ee^{itX}| \leqslant E|e^{itX}| = E1 = 1.$$

A razão pela qual funções características são finitas para todo t, enquanto funções geratrizes de momento, não são finitas em geral é que e^{it}, $-\infty < t < \infty$, é limitada, enquanto e^t, $-\infty < t < \infty$, não é limitada.

Exemplo 4. Seja X uma variável aleatória que assume o valor a com a probabilidade 1. Então

$$\varphi_X(t) = Ee^{itX} = e^{ita}, \qquad -\infty < t < \infty.$$

Em particular, se X assume o valor zero com probabilidade 1, sua função característica é identicamente igual a 1.

Se X é uma variável aleatória e, a e b são constantes reais, então

$$\begin{aligned}\varphi_{a+bX}(t) &= Ee^{it(a+bX)} \\ &= Ee^{ita}e^{ibtX} \\ &= e^{ita}Ee^{ibtX},\end{aligned}$$

de modo que

$$(10) \qquad \varphi_{a+bX}(t) = e^{ita}\varphi_X(bt), \qquad -\infty < t < \infty.$$

Exemplo 5. Suponha que U se distribui uniformemente em $(-1, 1)$. Então para $t \neq 0$

$$\begin{aligned}\varphi_U(t) &= \int_{-1}^{1} e^{itu}\frac{1}{2}\,du \\ &= \frac{1}{2}\frac{e^{itu}}{it}\Big|_{-1}^{1} \\ &= \frac{1}{2}\left(\frac{e^{it} - e^{-it}}{it}\right) \\ &= \frac{\sin t}{t}.\end{aligned}$$

Para $a < b$ seja

$$X = \frac{a+b}{2} + \left(\frac{b-a}{2}\right) U.$$

Então X distribui-se uniformemente em (a, b), e de acordo com (10) para $t \neq 0$

$$\varphi_X(t) = e^{it(a+b)/2} \frac{\sin((b-a)t/2)}{(b-a)t/2}.$$

Alternativamente

$$\varphi_X(t) = \int_a^b e^{itx} \frac{1}{b-a} \, dx$$

$$= \frac{1}{b-a} \left. \frac{e^{itx}}{it} \right|_a^b$$

$$= \frac{e^{ibt} - e^{iat}}{it(b-a)}.$$

É fácil checar por meio de (9) que estes dois resultados estão de acordo.

Exemplo 6. Suponha que X tem uma distribuição exponencial de parâmetro λ. Então

$$\varphi_X(t) = \int_0^\infty e^{itx} \lambda e^{-\lambda x} \, dx$$

$$= \lambda \int_0^\infty e^{-(\lambda - it)x} \, dx$$

$$= \frac{\lambda}{\lambda - it} \left. e^{-(\lambda - it)x} \right|_\infty^0.$$

Como $\lim_{x \to \infty} e^{-\lambda x} = 0$ e e^{itx} é limitado em x, segue-se que

$$\lim_{x \to \infty} e^{-(\lambda - it)x} = \lim_{x \to \infty} e^{-\lambda x} e^{itx} = 0.$$

Assim

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}.$$

Suponha que X e Y são variáveis aleatórias independentes. Então e^{itX} e e^{itY} são também variáveis aleatórias independentes; conseqüentemente

$$\varphi_{X+Y}(t) = Ee^{it(X+Y)} = Ee^{itX} e^{itY} = Ee^{itX} Ee^{itY}$$

e portanto

(11) $$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t), \qquad -\infty < t < \infty.$$

Pode-se estender a fórmula (11) para mostrar que a função característica da soma de um número finito de variáveis aleatórias é o produto das funções características individuais.

Pode-se mostrar que $\varphi_X(t)$ é uma função contínua de t. Além do mais, se X tem momento finito de ordem n, então $\varphi_X^{(n)}(t)$ existe, é contínua em t, e pode ser calculada como

$$\varphi_X^{(n)}(t) = \frac{d^n}{dt^n} Ee^{itX} = E \frac{d^n}{dt^n} e^{itX} = E(iX)^n e^{itX}.$$

Em particular

(12) $$\varphi_X^{(n)}(0) = i^n EX^n.$$

Podemos tentar expandir $\varphi_X(t)$ em série de potência de acordo com a fórmula

(13) $$\varphi_X(t) = Ee^{itX} = E \sum_{n=0}^{\infty} \frac{(itX)^n}{n!} = \sum_{n=0}^{\infty} \frac{i^n EX^n}{n!} t^n.$$

Suponha que

$$M_X(t) = \sum_{n=0}^{\infty} \frac{EX^n}{n!} t^n$$

é finito em $-t_0 < t < t_0$ para algum número positivo t_0. Então (13) é também válido em $-t_0 < t < t_0$.

Exemplo 7. Suponha que X se distribui normalmente com média 0 e variância σ^2. Obtenha $\varphi_X(t)$.

Sabemos do Capítulo 7 que $EX^n = 0$ para qualquer número inteiro positivo ímpar n. Além do mais, se $n = 2k$ é um inteiro par, então

$$EX^n = EX^{2k} = \frac{\sigma^{2k}(2k)!}{2^k k!}.$$

Portanto

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{i^{2k} EX^{2k}}{(2k)!} t^{2k} = \sum_{k=0}^{\infty} \frac{(-\sigma^2 t^2/2)^k}{k!} = e^{-\sigma^2 t^2/2}.$$

De uma forma mais geral suponha que X se distribui normalmente com média μ e variância σ^2. Então, $Y = X - \mu$ distribui-se normalmente com média 0 e variância σ^2. Como $X = \mu + Y$, vemos da fórmula (10) e do Exemplo 7 que

(14) $$\varphi_X(t) = e^{it\mu} e^{-\sigma^2 t^2/2}, \quad -\infty < t < \infty.$$

Seja X uma variável aleatória cuja função geratriz de momento é $M_X(t)$ finita em $-t_0 < t < t_0$ para algum número positivo t_0. Como

$$M_X(t) = Ee^{tX}$$

e

$$\varphi_X(t) = Ee^{itX},$$

deveríamos esperar que

$$\varphi_X(t) = M_X(it). \tag{15}$$

Em outras palavras, deveríamos esperar que substituindo t por it, na fórmula da função geratriz de momento, obteremos a fórmula da função característica correspondente. Isto realmente acontece, mas a compreensão total dos problemas envolvidos requer um conceito sofisticado (continuação analítica) da teoria de variável complexa.

Como um exemplo de (15), suponha que X se distribui normalmente com média μ e variância σ^2. Então como já vimos

$$M_X(t) = e^{\mu t} e^{\sigma^2 t^2/2}$$

e portanto

$$\begin{aligned} M_X(it) &= e^{\mu(it)} e^{\sigma^2 (it)^2/2} \\ &= e^{i\mu t} e^{-\sigma^2 t^2/2} \end{aligned}$$

que de acordo com (14) é $\varphi_X(t)$.

8.3. FÓRMULAS DE INVERSÃO E O TEOREMA DA CONTINUIDADE

Seja X uma variável aleatória inteira. Sua função característica é dada por

$$\varphi_X(t) = \sum_{-\infty}^{\infty} e^{ijt} f_X(j).$$

Uma das propriedades mais úteis de $\varphi_X(t)$ é que ela pode ser usada para calcular $f_X(k)$. Especificamente temos a "fórmula de inversão"

$$f_X(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \varphi_X(t)\, dt. \tag{16}$$

Para verificar (16) escrevemos o segundo membro desta fórmula como

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}\left[\sum_{-\infty}^{\infty} e^{ijt} f_X(j)\right] dt.$$

Um teorema da teoria de integração justifica a inversão da ordem de integração e somatório para obter a expressão

$$\sum_{-\infty}^{\infty} f_X(j)\,\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(j-k)t}\, dt.$$

Para completar a demonstração de (16) devemos mostrar que a última expressão é igual a $f_X(k)$. Para isso é suficiente mostrar que

$$\text{(17)} \qquad \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(j-k)t}\, dt = \begin{cases} 1 & \text{se } j = k, \\ 0 & \text{se } j \neq k. \end{cases}$$

A fórmula (17) é óbvia quando $j = k$, pois neste caso $e^{i(j-k)t} = 1$ para todo t. Se $j \neq k$, então

$$\begin{aligned} \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(j-k)t}\, dt &= \frac{e^{i(j-k)t}\big|_{-\pi}^{\pi}}{2\pi i(j-k)} \\ &= \frac{e^{i(j-k)\pi} - e^{-i(j-k)\pi}}{2\pi i(j-k)} \\ &= \frac{\operatorname{sen}(j-k)\pi}{\pi(j-k)} = 0, \end{aligned}$$

pois $\operatorname{sen} m\pi = 0$ para todos inteiros m. Isto completa a demonstração de (17) e portanto de (16).

Exemplo 8. Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias inteiras independentes e identicamente distribuídas e seja $S_n = X_1 + \cdots + X_n$. Então $\varphi_{S_n}(t) = (\varphi_{X_1}(t))^n$, e conseqüentemente de acordo com (16)

$$\text{(18)} \qquad f_{S_n}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}(\varphi_{X_1}(t))^n\, dt.$$

A fórmula (18) é a base para quase todos os métodos de analisar o comportamento de $f_{S_n}(k)$, para valores grandes de n e, em particular, a base para demonstração do Teorema de Limite Central "local" discutido no Capítulo 7.

Existe também um análogo de (16) para variáveis aleatórias contínuas. Seja X uma variável aleatória cuja função característica $\varphi_X(t)$ é integrável, isto é,

$$\int_{-\infty}^{\infty} |\varphi_X(t)|\, dt < \infty.$$

Pode-se mostrar que neste caso X é uma variável aleatória contínua cuja densidade f_X é dada por

$$(19) \qquad f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}\varphi_X(t)\,dt.$$

Exemplo 9. Suponha que X se distribui normalmente com média 0 e variância σ^2. Mostraremos diretamente que (19) é válido para tal variável aleatória. Do Exemplo 7 sabemos que a função característica de X é $\varphi_X(t) = e^{-\sigma^2 t^2/2}$. Assim pela definição de funções características,

$$e^{-\sigma^2 t^2/2} = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}\,dx.$$

Se substituimos t por $-t$ e σ por $1/\sigma$ nesta fórmula ela se transforma em

$$e^{-t^2/2\sigma^2} = \int_{-\infty}^{\infty} e^{-itx} \frac{\sigma}{\sqrt{2\pi}} e^{-\sigma^2 x^2/2}\,dx$$

ou equivalente

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-t^2/2\sigma^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}e^{-\sigma^2 x^2/2}\,dx.$$

Finalmente, invertendo as funções dos símbolos x e t na última equação obtemos

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}e^{-\sigma^2 t^2/2}\,dt,$$

que é simplesmente (19) neste caso especial.

Seja agora X uma variável aleatória qualquer. Seja Y uma variável aleatória independente de X que tem a distribuição normal padrão, e seja c uma constante positiva. Então $X + cY$ tem a função característica

$$\varphi_X(t)e^{-c^2t^2/2}.$$

Como $\varphi_X(t)$ é limitado em valor absoluto por 1 e $e^{-c^2t^2/2}$ é integrável, segue-se que $X + cY$ tem uma função característica integrável. Conseqüentemente (19) é aplicável e $X + cY$ é uma variável aleatória tendo uma densidade dada por

$$f_{X+cY}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi_X(t)e^{-c^2t^2/2}\,dt.$$

Se integramos ambos os lados desta equação sobre $a \leqslant x \leqslant b$ e invertemos a ordem de integração, concluímos que

$$P(a \le X + cY \le b) = \frac{1}{2\pi}\int_a^b \left(\int_{-\infty}^{\infty} e^{-itx}\varphi_X(t)e^{-c^2t^2/2}\,dt\right)dx$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_a^b e^{-itx}\,dx\right)\varphi_X(t)e^{-c^2t^2/2}\,dt$$

ou

(20) $$P(a \leq X + cY \leq b) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{e^{-ibt} - e^{-iat}}{-it} \right) \varphi_X(t) e^{-c^2t^2/2}\, dt.$$

A importância de (20) é que é válida para uma variável aleatória arbitrária X.

O segundo membro de (20) depende de X apenas através de $\varphi_X(t)$. Usando este fato e fazendo $c \to 0$ em (20), pode-se mostrar que a função de distribuição de X é determinada pela sua função característica. Este resultado é conhecido como "teorema de unicidade" e pode ser enunciado como segue:

Teorema 1. Se duas variáveis aleatórias têm a mesma função característica, elas têm a mesma função de distribuição.

Exemplo 10. Use o teorema de unicidade para mostrar que a soma de duas variáveis aleatórias independentes normalmente distribuídas têm distribuição normal.

Sejam X e Y variáveis aleatórias independentes distribuídas segundo $n(\mu_1, \sigma_1^2)$ e $n(\mu_2, \sigma_2^2)$, respectivamente. Então

$$\varphi_X(t) = e^{i\mu_1 t} e^{-\sigma_1{}^2 t^2/2}$$

e

$$\varphi_Y(t) = e^{i\mu_2 t} e^{-\sigma_2{}^2 t^2/2}.$$

Conseqüentemente

$$\varphi_{X+Y}(t) = e^{i(\mu_1 + \mu_2)t} e^{-(\sigma_1{}^2 + \sigma_2{}^2)t^2/2}.$$

Assim a função característica de $X + Y$ é a função característica de uma variável aleatória com distribuição normal de média $\mu_1 + \mu_2$ e variância $\sigma_1^2 + \sigma_2^2$. Pelo teorema de unicidade esta deve ser a distribuição de $X + Y$.

A aplicação mais importante da fórmula de inversão (20) é que ela pode ser usada para derivar o resultado seguinte, que é a base da demonstração da Lei Fraca dos Grandes Números e do Teorema do Limite Central.

Teorema 2. Sejam X_n, $n \geqslant 1$, e X variáveis aleatórias tais que

(21) $$\lim_{n \to \infty} \varphi_{X_n}(t) = \varphi_X(t), \qquad -\infty < t < \infty.$$

Então

(22) $$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

em todos os pontos x onde F_X é contínua.

Este teorema estabelece que a convergência das funções características implica na convergência das funções de distribuição correspondentes ou, em outras palavras,

que as funções de distribuição "dependem continuamente" de suas funções características. Por esta razão o Teorema 2 é conhecido comumente como "Teorema da Continuidade".

A demonstração deste teorema é razoavelmente complicada. Não apresentaremos os detalhes da demonstração, mas indicaremos brevemente as idéias principais de um método de demonstração.

Escolhemos em primeiro lugar uma variável aleatória Y que tem distribuição normal padrão e é independente das variáveis aleatórias X_n, $n \geq 1$.

Seja $a < b$ e seja c uma constante positiva. Então pela fórmula de inversão (20)

$$(23) \qquad P(a \leq X_n + cY \leq b) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{e^{-ibt} - e^{-iat}}{-it} \right) \varphi_{X_n}(t) e^{-c^2t^2/2} \, dt$$

e

$$(24) \qquad P(a \leq X + cY \leq b) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\frac{e^{-ibt} - e^{-iat}}{-it} \right) \varphi_X(t) e^{-c^2t^2/2} \, dt.$$

Por hipótese, $\varphi_{X_n}(t) \to \varphi_X(t)$ à medida que $n \to \infty$. Por um teorema da teoria da integração, segue-se que o segundo membro de (23) converge para o segundo membro de (24) à medida que $n \to \infty$. Conseqüentemente

$$(25) \qquad \lim_{n \to \infty} P(a \leq X_n + cY \leq b) = P(a \leq X + cY \leq b).$$

Existem mais duas etapas para a demonstração do teorema. Primeiro mostrar (fazendo $a \to -\infty$ em (25)) que

$$(26) \qquad \lim_{n \to \infty} P(X_n + cY \leq b) = P(X + cY \leq b).$$

Finalmente mostrar (fazendo $c \to 0$ em (26)) que

$$\lim_{n \to \infty} P(X_n \leq b) = P(X \leq b)$$

Sempre que $P(X = b) = 0$. O último resultado equivale à conclusão do teorema.

8.4. A LEI FRACA DOS GRANDES NÚMEROS E O TEOREMA DO LIMITE CENTRAL

Usaremos nesta seção o Teorema da Continuidade para demonstrar os dois importantes teoremas da teoria da probabilidade mencionados no título desta seção. Ambos os teoremas foram discutidos sem demonstração em capítulos anteriores. Para demonstrar estes teoremas precisamos inicialmente estudar o comportamento assintótico de $\log \varphi_X(t)$ próximo de $t = 0$.

Seja z um número complexo tal que $|z-1|<1$. Podemos definir $\log z$ através da série

$$\log z = (z-1) - \frac{(z-1)^2}{2} + \frac{(z-1)^3}{3} - \cdots$$

(outras definições de $\log z$ são necessárias para $|z-1| \geqslant 1$). Com esta definição temos as propriedades usuais de que $\log 1 = 0$,

$$e^{\log z} = z, \qquad |z-1| < 1,$$

e se $h(t)$, $a<t<b$, é uma função complexa diferenciável tal que $|h(t)-1|<1$, então

$$\frac{d}{dt}\log h(t) = \frac{h'(t)}{h(t)}.$$

Seja X uma variável aleatória tendo função característica $\varphi_X(t)$. Então $\varphi_X(t)$ é contínua e $\varphi_X(0)=1$. Assim $\log \varphi_X(t)$ é bem definido para t próximo de 0 e $\log \varphi_X(0) = 0$.

Suponha agora que X tem média finita μ. Então $\varphi_X(t)$ é diferenciável e, de acordo com (12), $\varphi_X'(0) = i\mu$. Portanto

$$\begin{aligned}
\lim_{t\to 0}\frac{\log \varphi_X(t)}{t} &= \lim_{t\to 0}\frac{\log \varphi_X(t) - \log \varphi_X(0)}{t-0} \\
&= \frac{d}{dt}\log \varphi_X(t)\Big|_{t=0} \\
&= \frac{\varphi_X'(0)}{\varphi_X(0)} \\
&= i\mu.
\end{aligned}$$

Conseqüentemente,

$$\lim_{t\to 0}\frac{\log \varphi_X(t) - i\mu t}{t} = 0. \tag{27}$$

Suponha a seguir que X também tem variância finita σ^2. Então $\varphi_X(t)$ é diferenciável duas vezes e de (12)

$$\varphi_X''(0) = -EX^2 = -(\mu^2 + \sigma^2).$$

Podemos aplicar a regra de l'Hôspital para obter

$$\lim_{t\to 0} \frac{\log \varphi_X(t) - i\mu t}{t^2} = \lim_{t\to 0} \frac{\dfrac{\varphi'_X(t)}{\varphi_X(t)} - i\mu}{2t}$$

$$= \lim_{t\to 0} \frac{\varphi'_X(t) - i\mu\varphi_X(t)}{2t\varphi_X(t)}$$

$$= \lim_{t\to 0} \frac{\varphi'_X(t) - i\mu\varphi_X(t)}{2t}.$$

Uma segunda aplicação da regra de l'Hôspital mostra que

$$\lim_{t\to 0} \frac{\log \varphi_X(t) - i\mu t}{t^2} = \frac{\varphi''_X(0) - i\mu\varphi'_X(0)}{2}$$

$$= \frac{-(\mu^2 + \sigma^2) - (i\mu)^2}{2}$$

$$= \frac{-\mu^2 - \sigma^2 + \mu^2}{2}.$$

Em outras palavras

(28) $$\lim_{t\to 0} \frac{\log \varphi_X(t) - i\mu t}{t^2} = -\frac{\sigma^2}{2}.$$

Teorema 3. Lei Fraca dos Grandes Números. Sejam $X_1, X_2, \ldots$ variáveis aleatórias independentes, identicamente distribuídas tendo média finita μ e seja $S_n = X_1 + \cdots + X_n$. Então para qualquer $\epsilon > 0$

(29) $$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0.$$

Demonstração. A função característica de

$$\frac{S_n}{n} - \mu = \frac{X_1 + \cdots + X_n}{n} - \mu$$

é

$$e^{-i\mu t}(\varphi_{X_1}(t/n))^n.$$

Suponha que t é fixo. Então para n suficientemente grande, t/n está próximo de zero o bastante para que $\log \varphi_{X_1}(t/n)$ seja bem definido e

(30) $$e^{-i\mu t}(\varphi_{X_1}(t/n))^n = \exp\left[n(\log \varphi_{X_1}(t/n) - i\mu(t/n))\right].$$

A seguir afirmamos que

(31) $$\lim_{n\to\infty} n(\log \varphi_{X_1}(t/n) - i\mu(t/n)) = 0.$$

A Equação (31) é óbvia para $t = 0$ uma vez que $\log \varphi_{X_1}(0) = \log 1 = 0$. Se $t \neq 0$ podemos escrever o primeiro membro de (31) como

$$t \lim_{n \to \infty} \frac{\log \varphi_X(t/n) - i\mu(t/n)}{t/n}.$$

Mas $t/n \to 0$ à medida que $n \to \infty$, de modo que o último limite é 0 de acordo com (27). Isto completa a demonstração de (31). Segue-se de (30) e (31) que a função característica de

$$\frac{S_n}{n} - \mu$$

se aproxima de 1 à medida que $n \to \infty$. Mas 1 é a função característica de uma variável aleatória X tal que $P(X = 0) = 1$. A função de distribuição de X é dada por

$$F_X(x) = \begin{cases} 1 & \text{se } x \geq 0, \\ 0 & \text{se } x < 0. \end{cases}$$

Esta função de distribuição é sempre contínua exceto para $x = 0$. Seja $\epsilon > 0$. Pelo Teorema da Continuidade,

$$\lim_{n \to \infty} P\left(\frac{S_n}{n} - \mu \leq -\varepsilon\right) = F_X(-\varepsilon) = 0 \tag{32}$$

e

$$\lim_{n \to \infty} P\left(\frac{S_n}{n} - \mu \leq \varepsilon\right) = F_X(\varepsilon) = 1.$$

O último resultado implica que

$$\lim_{n \to \infty} P\left(\frac{S_n}{n} - \mu > \varepsilon\right) = 0,$$

que juntamente com (32) mostra que (29) é válido como desejado.

Para o teorema seguinte é necessário lembrar que $\Phi(x)$ representa a função de distribuição normal padrão dada por

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy, \qquad -\infty < x < \infty.$$

Lembramos também que esta função de distribuição é contínua para todos os valores de x.

Teorema 4. Teorema do Limite Central. Sejam $X_1, X_2, \ldots$ variáveis independentes, identicamente distribuídas tendo média μ e variância finita não-nula σ^2. Então

$$\lim_{n \to \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), \qquad -\infty < x < \infty.$$

Demonstração. Seja

$$S_n^* = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Então para t fixo e n suficientemente grande,

$$\begin{aligned}\varphi_{S_n^*}(t) &= e^{-in\mu t/\sigma\sqrt{n}}\varphi_{S_n}(t/\sigma\sqrt{n}) \\ &= e^{-in\mu t/\sigma\sqrt{n}}(\varphi_{X_1}(t/\sigma\sqrt{n}))^n,\end{aligned}$$

ou

$$\varphi_{S_n^*}(t) = \exp\left[n(\log \varphi_{X_1}(t/\sigma\sqrt{n}) - i\mu(t/\sigma\sqrt{n}))\right]. \tag{33}$$

A seguir afirmamos que

$$\lim_{n\to\infty} n(\log \varphi_{X_1}(t/\sigma\sqrt{n}) - i\mu(t/\sigma\sqrt{n})) = -\frac{t^2}{2}. \tag{34}$$

Se $t = 0$, ambos os membros de (34) são iguais a zero e (34) é obviamente válido. Se $t \neq 0$ podemos escrever o primeiro membro de (34) como

$$\frac{t^2}{\sigma^2}\lim_{n\to\infty}\frac{\log \varphi_{X_1}(t/\sigma\sqrt{n}) - i\mu(t/\sigma\sqrt{n})}{(t/\sigma\sqrt{n})^2},$$

que de acordo com (28) é igual a

$$\frac{t^2}{\sigma^2}\left(-\frac{\sigma^2}{2}\right) = -\frac{t^2}{2}.$$

Assim (34) é válido para todo t. Segue-se de (33) e (34) que

$$\lim_{n\to\infty} \varphi_{S_n^*}(t) = e^{-t^2/2}, \qquad -\infty < t < \infty.$$

De acordo com o Exemplo 7 $e^{-t^2/2}$ é a função característica de uma variável aleatória X tendo a função de distribuição normal padrão $\Phi(x)$. Assim pelo Teorema de Continuidade.

$$\lim_{n\to\infty} P(S_n^* \leq x) = \Phi(x), \qquad -\infty < x < \infty,$$

que é a conclusão desejada.

Exercícios

1. Suponha que X se distribui uniformemente em (a, b). Obtenha $M_X(t)$.
2. Expresse a função geratriz de momento de $Y = a + bX$ em termos de $M_X(t)$ (onde a e b são constantes).
3. Suponha que X tem distribuição de Poisson de parâmetro λ. Use a função geratriz de momento para obter a média e variância de X.

4. Suponha que X tem distribuição binomial negativa de parâmetros α e p.

(a) Obtenha a função geratriz de momento de X.

(b) Use esta função geratriz de momento para obter a média e variância de X.

5. Seja X uma variável aleatória contínua tendo a densidade $f_X(x) = (1/2)e^{-|x|}$ $-\infty < x < \infty$.

(a) Mostre que $M_X(t) = 1/(1 - t^2)$, $-1 < t < 1$.

(b) Use esta função geratriz de momento para obter uma fórmula para EX^{2n} (observe que os momentos ímpares de X são todos zero).

6. Suponha que X tem distribuição binomial de parâmetros n e p.

(a) Obtenha $dM_X(t)/dt$ e $d^2M_X(t)/dt^2$.

(b) Use (*a*) e a fórmula (7) para obter a média e variância de X.

7. Sejam $X_1, \ldots, X_n$ variáveis aleatórias independentes identicamente distribuídas tais que $M_{X_1}(t)$ é finito para todo t. Use funções geratrizes de momento para mostrar que

$$E(X_1 + \cdots + X_n)^3 = nEX_1^3 + 3n(n-1)EX_1^2EX_1 + n(n-1)(n-2)(EX_1)^3.$$

Sugestão: obtenha $(d^3/dt^3)(M_{X_1}(t))^n|_{t=0}$.

8. Seja X uma variável aleatória tal que $M_X(t)$ é finito para todo t. Use o mesmo argumento da demonstração da Desigualdade de Chebyshev para concluir que

$$P(X \geqslant x) \leqslant e^{-tx}M_X(t), \qquad t \geqslant 0.$$

Segue-se que

$$P(X \geq x) \leq \min_{t \geq 0} e^{-tx}M_X(t),$$

desde que $e^{-tx}M_X(t)$ tenha um mínimo em $0 \leqslant t < \infty$.

9. Suponha que X tem distribuição gama de parâmetros α e λ. Use o resultado do Exercício 8 para mostrar que $P(X \geqslant 2\alpha/\lambda) \leqslant (2/e)^\alpha$.

10. Suponha que X tem distribuição de Poisson de parâmetro λ. Obtenha $\varphi_X(t)$.

11. Suponha que X tem distribuição geométrica de parâmetro p. Obtenha $\varphi_X(t)$.

12. Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes, cada uma com distribuição geométrica de parâmetro p. Obtenha a função característica de $X = X_1 + \cdots + X_n$.

13. Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias independentes, cada uma com distribuição exponencial de parâmetro λ. Obtenha a função característica de $X = X_1 + \cdots + X_n$.

14. Seja X uma variável aleatória discreta cujos valores possíveis são todos inteiros não-negativos. Que relação deveríamos esperar que exista entre a função característica de X e a função geratriz de probabilidade de X (lembre-se das fórmulas (3) e (15))?

15. Seja X uma variável aleatória qualquer.

    (a) Mostre que $\varphi_X(t) = E \cos tX + i\, E \operatorname{sen} tX$.

    (b) Mostre que $\varphi_{-X}(t) = E \cos tX - i\, E \operatorname{sen} tX$.

    (c) Mostre que $\varphi_{-X}(t) = \varphi_X(-t)$.

16. Seja X uma variável aleatória simétrica, isto é, tal que X e $-X$ têm a mesma função de distribuição.

    (a) Mostre que $E \operatorname{sen} tX = 0$ e que $\varphi_X(t)$ é real.

    (b) Mostre que $\varphi_X(-t) = \varphi_X(t)$.

17. Sejam X e Y variáveis aleatórias independentes e identicamente distribuídas. Mostre que $\varphi_{X-Y}(t) = |\varphi_X(t)|^2$. *Sugestão:* use o Exercício 15.

18. Seja X uma variável aleatória tal que $\varphi_X(t)$ é real.

    (a) Mostre que X e $-X$ têm a mesma função característica (use o Exercício 15)

    (b) Por que se segue que X e $-X$ têm a mesma função de distribuição?

19. Seja X uma variável aleatória contínua tendo a densidade $f_X(x) = (1/2)e^{-|x|}$, $-\infty < x < \infty$.

    (a) Mostre que $\varphi_X(t) = 1/(1 + t^2)$.

    (b) Use (a) e a fórmula de inversão (19) para concluir que

    $$e^{-|x|} = \int_{-\infty}^{\infty} e^{-ixt} \frac{1}{\pi(1 + t^2)}\, dt.$$

    (c) Mostre usando (b) que

    $$e^{-|x|} = \int_{-\infty}^{\infty} e^{ixt} \frac{1}{\pi(1 + t^2)}\, dt.$$

20. Seja X uma variável aleatória tendo a densidade de Cauchy

    $$f_X(x) = \frac{1}{\pi(1 + x^2)}, \qquad -\infty < x < \infty.$$

    Mostre que $\varphi_X(t) = e^{-|t|}$, $-\infty < t \infty$. *Sugestão:* inverta as funções de x e t no Exercício 19.

21. Sejam X e Y variáveis aleatórias independentes tendo a densidade de Cauchy.

    (a) Obtenha as funções características de $X + Y$ e de $(X + Y)/2$.

    (b) Por que se segue que $(X + Y)/2$ também tem a densidade de Cauchy?

22. Estenda o resultado do Exercício 21 mostrando que se $X_1, X_2, \ldots, X_n$ são variáveis aleatórias independentes, cada uma com densidade de Cauchy, então $(X_1 + \cdots + X_n)/n$ também tem a densidade de Cauchy.

23. Para $\lambda > 0$ seja X_λ uma variável aleatória tendo distribuição de Poisson de parâmetro λ.

(a) Usando argumentos semelhantes àqueles usados na demonstração do Teorema do Limite Central mostre que para $-\infty < t < \infty$,

$$\lim_{\lambda \to \infty} Ee^{it(X_\lambda - \lambda)/\sqrt{\lambda}} = \lim_{\lambda \to \infty} \exp \left[\lambda(e^{it/\sqrt{\lambda}} - 1 - it/\sqrt{\lambda})\right] = e^{-t^2/2}.$$

(b) Que conclusão se deve tirar de (a), através de uma modificação adequada do Teorema da Continuidade?

9

CAMINHOS ALEATÓRIOS E PROCESSOS DE POISSON

Neste capítulo discutimos dois exemplos elementares porém importantes de processos estocásticos. Um *processo estocástico* pode ser definido como qualquer coleção de variáveis aleatórias. Geralmente, entretanto, quando nos referimos a um processo estocástico temos em mente um processo que possua estrutura adicional suficiente para que resultados interessantes e úteis possam ser derivados. Isto certamente é verdade para os dois exemplos tratados neste capítulo. O material sobre processos de Poisson, nosso segundo exemplo, não depende das duas primeiras seções onde discutimos caminhos aleatórios.

9.1. CAMINHOS ALEATÓRIOS

Considere uma seqüência de jogos tal que, durante o n-ésimo jogo, observa-se uma variável aleatória X_n e qualquer jogador participando do n-ésimo jogo, recebe a importância X_n da "banca" (naturalmente o jogador paga na realidade $-X_n$ à banca se $X_n < 0$).

Acompanhemos o processo de um jogador que começa com um capital inicial x. Seja $S_n, n \geqslant 0$, o seu capital após n jogos. Então $S_0 = x$ e

$$S_n = x + X_1 + \cdots + X_n, \qquad n \geqslant 1.$$

A coleção de variáveis aleatórias $S_0, S_1, S_2, \ldots$ é um exemplo de um processo estocástico. Para obter resultados interessantes admitiremos a hipótese de que as variáveis aleatórias $X_1, X_2, \ldots$ são independentes e identicamente distribuídas. Sob essa hipótese o processo $S_0, S_1, \ldots$ chama-se *caminho aleatório.*

Admitiremos a hipótese adicional de que os X_k têm média finita μ. Se um jogador participa dos n primeiros jogos, seu capital esperado no momento da conclusão do n-ésimo jogo é

$$ES_n = x + n\mu. \tag{1}$$

Suponha entretanto que o jogador escolha os números $a \leqslant x$ e $b \geqslant x$ e assuma consigo o compromisso de abandonar o jogo quando seu capital se tornar não maior que a ou não menor que b. Então o número T de vezes que ele joga é uma variável aleatória definida por

$$T = \min(n \geqslant 0 \mid S_n \leqslant a \text{ ou } S_n \geqslant b).$$

Para garantir que $S_n \leqslant a$ ou $S_n \geqslant b$ para algum n, admitimos que

$$P(X_k = 0) < 1. \tag{2}$$

É possível demonstrar que a variável aleatória T é finita (com probabilidade 1) e na realidade $P(T > n)$ decresce exponencialmente à medida que $n \to \infty$. Isto significa que existem constantes positivas M e $c < 1$ tais que

$$P(T > n) < Mc^n, \quad n = 0, 1, 2, \ldots \tag{3}$$

A demonstração de (3) não é difícil mas será omitida para reservar espaço para resultados de interesse muito maior. De (3) e do Teorema 5 do Capítulo 4, segue-se que ET e todos os momentos de ordem superior de T são finitos.

Se o jogador abandona o jogo após a T-ésima aposta, seu capital será S_T (ver Figura 1). Uma famosa identidade devida a Abraham Wald relaciona o capital esperando, quando o jogador abandona o jogo, com o número esperado

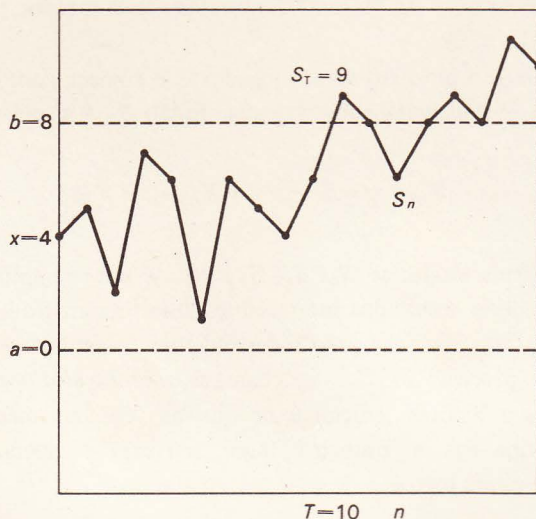



Figura 1

de vezes que ele participa do jogo. Especificamente, a identidade de Wald estabelece que

(4) $$ES_T = x + \mu ET.$$

A identidade de Wald é extraordinariamente semelhante a (1).

É conveniente introduzir uma nova notação para demonstrar a identidade de Wald. Seja A um evento qualquer. Representaremos por 1_A a variável aleatória indicadora de A, isto é, a variável aleatória que é 1 se A ocorre e, 0 se A não ocorre. Por definição $1_A + 1_{A^c} = 1$. Usando esta notação podemos escrever

$$S_T = x + \sum_{j=1}^{T} X_j = x + \sum_{j=1}^{\infty} X_j 1_{\{T \geq j\}}.$$

Uma vez que o complemento do evento $\{T \geqslant j\}$ é o evento $\{T < j\}$, vemos que

(5) $$S_T = x + \sum_{j=1}^{\infty} X_j(1 - 1_{\{T<j\}}),$$

e portanto

(6) $$ES_T = x + E \sum_{j=1}^{\infty} X_j(1 - 1_{\{T<j\}}).$$

Pode-se mostrar usando a teoria da media que a ordem de expectância e somatório pode ser invertida em (6). Assim

(7) $$ES_T = x + \sum_{j=1}^{\infty} E[X_j(1 - 1_{\{T<j\}})].$$

Para determinar se $T<j$, não basta olhar as variáveis aleatórias $X_1, X_2, \ldots, X_{j-1}$. Segue-se que as variáveis aleatórias X_j e $(1 - 1_{\{T<j\}})$ são independentes. Conseqüentemente

$$\begin{aligned} E[X_j(1 - 1_{\{T<j\}})] &= EX_j E(1 - 1_{\{T<j\}}) \\ &= \mu(1 - P(T < j)) \\ &= \mu P(T \geq j). \end{aligned}$$

Concluímos de (6) e do Teorema 5 do Capítulo 4 que

$$\begin{aligned} ES_T &= x + \mu \sum_{j=1}^{\infty} P(T \geq j) \\ &= x + \mu ET, \end{aligned}$$

que completa a demonstração da identidade de Wald.

Se os X_n têm média $\mu = 0$ e variância finita σ^2, existe uma segunda identidade devida, a Wald, a saber

(8) $$\text{Var } S_T = \sigma^2 ET.$$

Como $ES_T = x$ em (4), a fórmula (8) é equivalente a

$$(9) \qquad E(S_T - x)^2 = \sigma^2 ET.$$

A seguir verificaremos (9). De acordo com (5)

$$S_T - x = \sum_{j=1}^{\infty} X_j(1 - 1_{\{T<j\}}),$$

e portanto

$$(S_T - x)^2 = \sum_{j=1}^{\infty} X_j(1 - 1_{\{T<j\}}) \sum_{k=1}^{\infty} X_k(1 - 1_{\{T<k\}})$$

$$= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} X_j(1 - 1_{\{T<j\}})X_k(1 - 1_{\{T<k\}}).$$

Para tomar as expectâncias é novamente permissível inverter expectância e somatório. Concluímos que

$$(10) \qquad E(S_T - x)^2 = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} E[X_j(1 - 1_{\{T<j\}})X_k(1 - 1_{\{T<k\}})].$$

Avaliaremos a seguir os termos individuais deste somatório duplo. Considere inicialmente os termos correspondentes a valores de j e k tais que $j < k$. A variável aleatória

$$X_j(1 - 1_{\{T<j\}})(1 - 1_{\{T<k\}})$$

depende somente de $X_1, X_2, \ldots, X_{k-1}$, e portanto, é independente da variável aleatória X_k. Como $EX_k = \mu = 0$, vemos que

$$E[X_j(1 - 1_{\{T<j\}})X_k(1 - 1_{\{T<k\}})]$$
$$= E[X_j(1 - 1_{\{T<j\}})(1 - 1_{\{T<k\}})]EX_k = 0.$$

Analogamente os termos do segundo membro de (10) desaparecem quando $j > k$. Quando $j = k$ obtemos

$$E[X_j^2(1 - 1_{\{T<j\}})^2].$$

A variável aleatória $(1 - 1_{\{T<j\}})$ depende somente de $X_1, X_2, \ldots, X_{j-1}$, e portanto é independente de X_j. Como esta variável aleatória assume apenas os valores 0 e 1, vemos que

$$(1 - 1_{\{T<j\}})^2 = 1 - 1_{\{T<j\}}.$$

Assim

$$E[X_j^2(1 - 1_{\{T<j\}})^2] = E[X_j^2(1 - 1_{\{T<j\}})]$$
$$= EX_j^2 E(1 - 1_{\{T<j\}})$$
$$= \text{Var}\, X_j^2(1 - P(T < j))$$
$$= \sigma^2 P(T \geq j).$$

Concluímos de (10) e do Teorema 5 do Capítulo 4 que

$$E(S_T - x)^2 = \sigma^2 \sum_{j=1}^{\infty} P(T \geq j) = \sigma^2 ET,$$

o que demonstra (9) e portanto (8).

9.2. CAMINHOS ALEATÓRIOS SIMPLES

Admitiremos ao longo de toda esta seção que $a \leqslant x \leqslant b$, $a < b$ e que a, b e x são inteiros. Pode-se aplicar mais facilmente as identidades da seção anterior se sabemos que

$$(11) \qquad P(S_T = a \text{ ou } b) = P(S_T = a) + P(S_T = b) = 1.$$

Este certamente será o caso de

$$(12) \qquad P(X_k = -1, 0, \text{ ou } 1) = 1,$$

e neste caso o caminho aleatório $S_0, S_1, \ldots$ chama-se *caminho aleatório simples.* A propriedade principal que distingue caminhos aleatórios simples de outros caminhos aleatórios é que eles não dão "saltos sobre" pontos inteiros. Sejam

$$p = P\{X_k = 1\},$$
$$q = P\{X_k = -1\},$$
$$r = P\{X_k = 0\}.$$

Então $p \geqslant 0$, $q \geqslant 0$, $r \geqslant 0$ e $p + q + r = 1$. A hipótese (2) estabelece que $r < 1$. Segue-se de (11) que

$$(13) \qquad \begin{aligned} ES_T &= aP(S_T = a) + bP(S_T = b) \\ &= a(1 - P(S_T = b)) + bP(S_T = b). \end{aligned}$$

Para caminhos aleatórios simples podemos resolver explicitamente para $P(S_T = a)$, $P(S_T = b)$, ES_T e ET. Considere inicialmente o caso $p = q$. Então $\mu = 0$ e a identidade de Wald (4) se transforma em $ES_T = x$. Assim como (13)

$$x = a(1 - P(S_T = b)) + bP(S_T = b).$$

Segue-se imediatamente que

$$(14) \qquad P(S_T = b) = \frac{x - a}{b - a}$$

e

$$(15) \qquad P(S_T = a) = \frac{b - x}{b - a}.$$

A identidade de Wald não dá informação alguma sobre ET quando $\mu = 0$. A identidade (8) é aplicável neste caso e obtemos que $\sigma^2 = p + q = 1 - r$ e

$$\text{Var } S_T = \sigma^2 ET = (1 - r)ET.$$

Mas

$$\begin{aligned}\text{Var } S_T &= ES_T^2 - (ES_T)^2 \\ &= b^2 P(S_T = b) + a^2 P(S_T = a) - x^2 \\ &= \frac{b^2(x - a) + a^2(b - x)}{b - a} - x^2 \\ &= (ax + bx - ab) - x^2 \\ &= (x - a)(b - x).\end{aligned}$$

Assim se $p = q$,

$$ET = \frac{(x - a)(b - x)}{1 - r}. \tag{16}$$

Se $r = 0$ e $p = q = 1/2$,

$$ET = (x - a)(b - x). \tag{17}$$

Exemplo 1. Dois jogadores tendo capitais iniciais de \$5 e \$10, respectivamente, concordam em fazer uma série de apostas até a ruína de um deles. Suponha que os resultados das apostas são independentes e que ambos os jogadores têm probabilidade 1/2 de ganhar qualquer aposta. Determine a probabilidade de que ocorra a ruína do jogador cujo capital inicial é \$10. Determine o número esperado de apostas.

O problema se encaixa no nosso esquema com S_n representando o capital do jogador menos rico, ao cabo de n apostas, se escolhemos $p = q = 1/2$, $x = 5$, $a = 0$, e $b = 15$. A resposta da primeira parte é

$$P(S_T = b) = \frac{5 - 0}{15 - 0} = \frac{1}{3}.$$

A resposta da segunda parte é

$$ET = (5 - 0)(15 - 5) = 50.$$

Suponha a seguir que $p \neq q$. Para evitar exceções triviais admitiremos que $p > 0$ e $q > 0$. A identidade de Wald não pode ser usada para obter $P(S_T = b)$ se $p \neq q$; portanto necessitamos de outra abordagem.

Defina $f(x)$ para x inteiro em $[a, b]$ fazendo $f(x)$ representar a probabilidade de $S_T = b$ quando $S_0 = x$. Observamos inicialmente que f satisfaz a equação de diferenças finitas

(18) $$f(x) = pf(x+1) + qf(x-1) + rf(x), \qquad a < x < b.$$

Isto é verdade porque

$$\begin{aligned} f(x) &= P(S_T = b) \\ &= p \cdot P(S_T = b \mid X_1 = 1) + q \cdot P(S_T = b \mid X_1 = -1) \\ &\quad + r \cdot P(S_T = b \mid X_1 = 0) \end{aligned}$$

e

$$P(S_T = b \mid X_1 = i) = f(x+i), \qquad i = 1, -1, 0.$$

Além de (18) temos as óbvias condições de contorno

(19) $$f(a) = 0 \quad \text{e} \quad f(b) = 1.$$

De (18) e $(1 - r) = p + q$, vemos que

(20) $$f(x+1) - f(x) = \frac{q}{p}\,(f(x) - f(x-1)), \qquad a < x < b.$$

Seja $c = f(a+1) = f(a+1) - f(a)$. Então (20) implica que

$$f(x+1) - f(x) = c\left(\frac{q}{p}\right)^{x-a}, \qquad a \leq x < b.$$

Usando a fórmula da soma de uma progressão geométrica obtemos

$$\begin{aligned} f(x) = f(x) - f(a) &= \sum_{y=a}^{x-1} (f(y+1) - f(y)) \\ &= \sum_{y=a}^{x-1} c\left(\frac{q}{p}\right)^{y-a} \\ &= c\,\frac{1 - (q/p)^{x-a}}{1 - (q/p)}, \qquad a \leq x \leq b. \end{aligned}$$

Do caso especial $f(b) = 1$, temos agora que

$$c = \frac{1 - (q/p)}{1 - (q/p)^{b-a}}.$$

Substituindo este resultado na expressão de $f(x)$, obtemos

$$f(x) = \frac{1 - (q/p)^{x-a}}{1 - (q/p)^{b-a}}.$$

Em outras palavras, mostramos que se $p \neq q$ e $p > 0$, então

$$(21) \qquad P(S_T = b) = \frac{1 - (q/p)^{x-a}}{1 - (q/p)^{b-a}}, \qquad a \leqslant x \leqslant b.$$

Segue-se imediatamente de (21) que sob as mesmas condições

$$(22) \qquad P(S_T = a) = \frac{(q/p)^{x-a} - (q/p)^{b-a}}{1 - (q/p)^{b-a}}, \qquad a \leqslant x \leqslant b.$$

De (13) e (21)

$$(23) \qquad ES_T = (b - a)\,\frac{1 - (q/p)^{x-a}}{1 - (q/p)^{b-a}} + a.$$

Como $\mu = p - q$, segue-se agora da identidade de Wald que

$$(24) \qquad ET = \left(\frac{b - a}{p - q}\right) \frac{1 - (q/p)^{x-a}}{1 - (q/p)^{b-a}} - \frac{x - a}{p - q}, \qquad a \leq x \leq b.$$

Exemplo 2. Modifiquemos o exemplo anterior supondo que o jogador mais rico, em virtude de sua maior habilidade, tem probabilidade 0,6 de ganhar uma aposta qualquer, com o outro jogador tendo probabilidade 0,4 de ganhar a aposta. Determine a probabilidade de que ocorra a ruína ao jogador mais rico, o ganho esperado deste jogador e o número esperado de apostas.

Para este caso tomamos $p = 0{,}4$ e $q = 0{,}6$. A probabilidade de que ocorra a ruína do jogador mais rico é

$$P(S_T = 15) = \frac{1 - (0{,}6/0{,}4)^5}{1 - (0{,}6/0{,}4)^{15}} = 0{,}0151.$$

Para obter o ganho esperado do jogador mais rico observamos inicialmente que o capital esperado do jogador mais pobre ao final do jogo é

$$ES_T = 15P(S_T = 15) = 15(0{,}0151) = 0{,}23.$$

Assim o ganho esperado do jogador mais rico ou a perda esperada do jogador mais pobre é \$5,00 − \$0,23 = \$4,77, uma boa percentagem do capital inicial do jogador mais pobre. O número esperado de apostas é

$$ET = \frac{ES_T - x}{\mu} = \frac{-4{,}77}{-0{,}2} \approx 24.$$

Suponha que $b \to \infty$ em (21). Pode-se mostrar que o primeiro membro de (21) converge para $P(S_n > a$ para todo $n \geqslant 0)$. Se $q < p$, o segundo membro de (21) converge para $1 - (q/p)^{x-a}$. Se $q > p$, o segundo membro de (21) converge para 0. Se $q = p$, o segundo membro de (14) converge para 0. Assim para $a \leqslant x = S_0$,

$$(25) \qquad P(S_n > a \text{ para todo } n \geq 0) = \begin{cases} 1 - \left(\dfrac{q}{p}\right)^{x-a} & \text{para } q < p, \\ 0 & \text{para } q \geq p. \end{cases}$$

De forma semelhante para $b \geqslant x = S_0$

$$P(S_n < b \text{ para todo } n \geq 0) = \begin{cases} 1 - \left(\dfrac{p}{q}\right)^{b-x} & \text{para } p < q, \\ 0 & \text{para } p \geq q. \end{cases}$$

Exemplo 3. Um cassino tem um capital de cem mil dólares. Um jogador infinitamente rico tenta quebrar o cassino. Uma vez aceito seu desafio, ele decide apostar \$1000 de cada vez. Se o jogador tem probabilidade 0,49 de ganhar cada aposta e 0,51 de perder, qual é a probabilidade de que ele consiga quebrar a casa?

Seja S_n o capital do cassino (em múltiplos de \$1000) após n jogos. Então $p = 0{,}51$, $q = 0{,}49$, $x = 100$ e $a = 0$. De acordo com (25) a probabilidade de que ocorra a ruína do cassino é

$$1 - P(S_n > 0 \text{ para todo } n \geqslant 0) = \left(\frac{q}{p}\right)^{x-a} = \left(\frac{0{,}49}{0{,}51}\right)^{100} = 0{,}018.$$

Seja A um subconjunto dos inteiros (nas aplicações A terá 0, 1, ou 2 pontos). Para $x \notin A$ e $y \notin A$, seja $P_A(x,y)$ a probabilidade de que um caminho aleatório simples partindo de x alcance y em algum tempo positivo antes de alcançar A. Para $x \in A$ ou $y \in A$, seja $P_A(x,y) = 0$. Estas probabilidades podem ser determinadas em termos das fórmulas desta seção.

Exemplo 4. Suponha que $p = q$. Determine $P_{a,b}(y,y)$, onde $a < y < b$.

Depois de um passo, o caminho aleatório está em $y - 1$, y, ou $y + 1$ com probabilidades p, $1 - 2p$ e p, respectivamente. De $y - 1$, a probabilidade de retornar a y antes de alcançar a é $(y - a - 1)/(y - a)$. De $y + 1$, a probabilidade de retornar a y antes de alcançar b é $(b - y - 1)/(b - y)$. Assim a probabilidade de retornar a y antes de alcançar a ou b é dada por

$$P_{\{a,b\}}(y,y) = p\,\frac{y - a - 1}{y - a} + 1 - 2p + p\,\frac{b - y - 1}{b - y}$$

ou

$$(26) \qquad P_{\{a,b\}}(y,y) = 1 - \frac{p(b-a)}{(y-a)(b-y)}.$$

Para $x \notin A$ e $y \notin A$, seja $G_A(x, y)$ o número esperado de visitas a y (para valores positivos de n) antes de atingir A para um caminho aleatório começando em x. Seja $G_A(x,y) = 0$ se $x \in A$ ou $y \in A$. Não é difícil mostrar que o número de retornos de y a y, antes de atingir A, tem distribuição geométrica de parâmetro $p = 1 - P_A(y, y)$. Assim pelo Exemplo 3 do Capítulo 4,

(27) $$G_A(y,y) = \frac{P_A(y,y)}{1 - P_A(y,y)}.$$

Se $x \neq y$, então

(28) $$G_A(x,y) = P_A(x,y)(1 + G_A(y,y)).$$

Pois para ter qualquer número positivo de visitas a y antes de atingir A, devemos primeiro chegar a y antes de atingir A.

Isto ocorre com probabilidade $P_A(x, y)$. Se chegamos a y, o número total de visitas a y antes de alcançar A é 1 mais o número total de retornos de y a y antes de alcançar A. Isto explica (28). De (27) e (28) temos que

(29) $$G_A(x,y) = \frac{P_A(x,y)}{1 - P_A(y,y)} \quad \text{para todo } x \text{ e } y.$$

Exemplo 5. Retornemos ao primeiro exemplo desta seção. Determine o número esperado de vezes $n > 1$ que os dois jogadores retornam aos seus capitais iniciais antes que ocorra a ruína de um deles.

Lembramos que neste exemplo $p = q = 1/2$, $a = 0$, $x = 5$ e $b = 15$. A probabilidade de retornar aos capitais originais antes que ocorra a ruína de um deles é, de acordo com (26),

$$P_{0,15}(5,5) = 1 - \frac{(1/2)(15)}{5 \cdot 10} = 0{,}85.$$

Assim de acordo com (27) o número esperado de vezes que ambos os jogadores voltam aos seus capitais iniciais antes que ocorra a ruína de um deles é

$$G_{0,15}(5,5) = \frac{P_{0,15}(5,5)}{1 - P_{0,15}(5,5)}$$
$$= \frac{0{,}85}{0{,}15} = 5{,}67.$$

9.3. CONSTRUÇÃO DE UM PROCESSO DE POISSON

Nas seções restantes deste capítulo consideraremos um modelo probabilístico para a distribuição aleatória de partículas no espaço ou eventos no tempo. Tais modelos encontram aplicações em uma variedade de campos diferentes. Como um exemplo, suponha que temos uma porção de uma substância radioativa e suponha que um experimento consiste em observar os tempos em que ocorrem as desintegrações. Suponha que o experimento começa no tempo 0 e seja D_m o tempo da

m-ésima desintegração. Como foi discutido no Exemplo 2 do Capítulo 1, as leis da física nos diz que os tempos D_m devem ser considerados variáveis aleatórias. A coleção de pontos $D_1, D_2, \ldots$ pode ser considerada como um subconjunto aleatório enumerável de $[0, \infty)$.

Como outra ilustração de um fenômeno que é essencialmente o mesmo, considere as chamadas chegando a uma central telefônica. Seja D_m o tempo em que inicia a m-ésima chamada. Não se conhece nenhuma forma de prever D_m com exatidão, mas elas podem ser tratadas como variáveis aleatórias para obtenção de resultados úteis.

Considere um experimento do seguinte tipo. Introduz-se um bastão em um tubo contendo uma suspensão de bactérias. A seguir esfrega-se o bastão uniformemente sobre a superfície de um disco contendo nutrientes através dos quais as bactérias podem se multiplicar. Após alguns dias aparece uma colônia visível de bactérias em cada lugar onde foi depositada uma bactéria. A localização destes pontos bem como o seu número total são aleatórios. A Figura 2 ilustra esta situação.

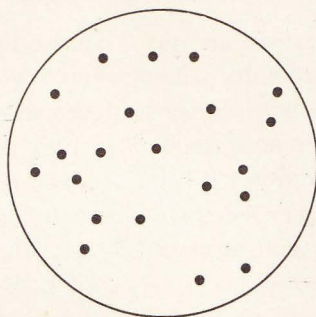



Figura 2

A localização dos pontos pode ser vista como um subconjunto aleatório dos pontos do disco.

Nos exemplos de partículas radioativas e bactérias fomos levados a considerar uma coleção aleatória de pontos em um certo subconjunto S do espaço Euclidiano. Nestes exemplos, tanto a localização das "partículas" como seu número total são aleatórios. Associadas a uma coleção aleatória deste tipo existem diversas variáveis aleatórias tais como o número total N de partículas em S, o número de partículas N_B em um dado subconjunto $B \subset S$ e a distância D_m de um dado ponto em S até a m-ésima partícula mais próxima. Na Figura 3 as partículas são representadas por pontos. Nesta figura $N = N_S = 4$ e $N_B = 2$. Estudaremos as distribuições e as distribuições conjuntas de variáveis aleatórias como estas no restante deste capítulo.

Existem naturalmente inúmeros modelos matemáticos diferentes que consideram a distribuição aleatória de partículas. Consideremos um dos mais elementares e mais importantes de tais modelos, chamado processo de Poisson. Tal processo

está intimamente relacionado com a distribuição uniforme de partículas que será considerada inicialmente.

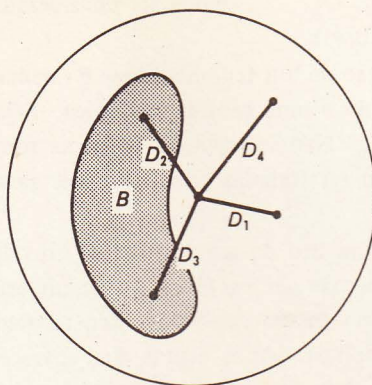



Figura 3

Considere então um sistema em que o número total de partículas n é fixo, mas para o qual as localizações das partículas em S são aleatórias. O modelo que desejamos é tal que estas n partículas se distribuem independente e uniformemente sobre o conjunto S tendo volume finito. Seja $|B|$ o volume de B. Então cada partícula tem probabilidade $p = |B|/|S|$ de cair em B. Conseqüentemente o número de partículas N_B que caem em B tem distribuição binomial de parâmetros n e p. De uma forma mais geral, sejam $B_1, B_2, \ldots, B_k$, k subconjuntos disjuntos de S cuja união é S e seja $p_j = |B_j|/|S|$. Então as variáveis aleatórias $N_{B_1}, \ldots, N_{B_k}$ tem distribuição multinomial de parâmetros n e $p_1, \ldots, p_k$. Portanto se $n_1, \ldots, n_k$ são números inteiros não-negativos de soma n,

$$P(N_{B_1} = n_1, \ldots, N_{B_k} = n_k) = \frac{n!}{(n_1!)\cdots(n_k!)} p_1^{n_1}\cdots p_k^{n_k}$$

$$= \frac{n!}{|S|^n} \prod_{j=1}^{k} \frac{|B_j|^{n_j}}{n_j!}.$$

O processo de Poisson em S resulta de uma modificação do sistema acima. Supomos agora que o número total de partículas $N = N_S$ em S, é uma variável aleatória tendo uma distribuição de Poisson de parâmetro $\lambda |S|$. Além do mais admitimos que dado $N = n$, as partículas se distribuem independente e uniformemente sobre S. Considere os subconjuntos $B_1, \ldots, B_k$ definidos anteriormente. Nossas hipóteses são de que se $n_1, \ldots, n_k$ são k números inteiros não-negativos de soma n, então

$$P(N_{B_1} = n_1, \ldots, N_{B_k} = n_k \mid N = n) = \frac{n!}{|S|^n} \prod_{j=1}^{k} \frac{|B_j|^{n_j}}{n_j!}.$$

Portanto

$$(30) \qquad P(N = n, N_{B_1} = n_1, \ldots, N_{B_k} = n_k) = P(N = n) \frac{n!}{|S|^n} \prod_{j=1}^{k} \frac{|B_j|^{n_j}}{n_j!}$$

$$= \frac{\lambda^n |S|^n}{n!} e^{-\lambda|S|} \frac{n!}{|S|^n} \prod_{j=1}^{k} \frac{|B_j|^{n_j}}{n_j!}$$

$$= \lambda^n e^{-\lambda|S|} \prod_{j=1}^{k} \frac{|B_j|^{n_j}}{n_j!}.$$

Como os conjuntos B_j são disjuntos e sua união é S,

$$|S| = |B_1| + \cdots + |B_k|,$$

de modo que o segundo membro de (30) pode ser escrito como

$$\prod_{j=1}^{k} \frac{(\lambda|B_j|)^{n_j}}{n_j!} e^{-\lambda|B_j|}.$$

O evento $\{N = n, N_{B_1} = n_1, \ldots, N_{B_k} = n_k\}$ é o mesmo que o evento $\{N_{B_1} = n_1, \ldots, N_{B_k} = n_k\}$ porque $n = n_1 + \cdots + n_k$ e $N = N_{B_1} + \cdots + N_{B_k}$. Portanto

$$P(N = n, N_{B_1} = n_1, \ldots, N_{B_k} = n_k) = P(N_{B_1} = n_1, \ldots, N_{B_k} = n_k).$$

Acabamos assim de demonstrar um importante fato:

$$(31) \qquad P(N_{B_1} = n_1, \ldots, N_{B_k} = n_k) = \prod_{j=1}^{k} \frac{(\lambda|B_j|)^{n_j}}{n_j!} e^{-\lambda|B_j|}.$$

Em outras palavras, *as variáveis aleatórias* $N_{B_1}, \ldots, N_{B_k}$ *são independentes e têm distribuições de Poisson de parâmetros* $\lambda|B_j|, \ldots, \lambda|B_k|$, *respectivamente.*

Não é muito surpreendente que as variáveis aleatórias N_{B_j}, $1 \leqslant j \leqslant k$, tenham distribuições de Poisson; mas é surpreendente que sejam mutuamente independentes porque no caso em que o número de partículas é fixo, as quantidades correspondentes são dependentes. É esta propriedade de independência que torna fácil o manuseio do processo de Poisson em aplicações.

Nos modelos acima, o número total de partículas, fosse ele fixo ou aleatório, era sempre finito, e as partículas se distribuíam sobre conjuntos tendo volume total finito. Entretanto para certas finalidades é teoricamente mais simples considerar um número infinito de partículas distribuídas sobre um conjunto tendo volume infinito. Assim poderíamos desejar que as partículas se distribuíssem sobre todo R^r ou todo $[0, \infty)$ etc. Para cobrir tais casos, necessitamos apenas introduzir uma pequena extensão no modelo anterior.

A hipótese básica de um processo de Poisson sobre um conjunto S *tendo volume finito ou infinito é que se* $B_1, \ldots, B_k$ *são subconjuntos disjuntos de* S

tendo volumes finitos, as variáveis aleatórias $N_{B_1}, \ldots, N_{B_k}$ *são independentes e têm distribuições de Poisson de parâmetros* $\lambda \mid B_1 \mid, \ldots, \lambda \mid B_k \mid$, *respectivamente.* A constante λ chama-se parâmetro do processo de Poisson.

Seja B um subconjunto de S tendo volume finito. Como uma conseqüência da definição de processo de Poisson, segue-se que se $B_1, B_2, \ldots, B_k$ são subconjuntos de B cuja união é B e $n_1, n_2, \ldots, n_k$ são números inteiros não-negativos cuja soma é n, então

$$(32) \qquad P(N_{B_1} = n_1, \ldots, N_{B_k} = n_k \mid N_B = n) = \frac{n!}{|B|^n} \prod_{j=1}^{k} \frac{|B_j|^{n_j}}{n_j!}.$$

Para verificar (32) observe que a probabilidade desejada é simplesmente

$$\frac{P(N_{B_1} = n_1, \ldots, N_{B_k} = n_k)}{P(N_B = n)} = \frac{\prod_{j=1}^{k} e^{-\lambda|B_j|} (\lambda|B_j|)^{n_j}/n_j!}{e^{-\lambda|B|} (\lambda|B|)^n/n!},$$

que se reduz ao segundo membro de (32).

Outra interpretação de (32) é a seguinte: Dado que existem n partículas em B, a distribuição conjunta das variáveis aleatórias $N_{B_1}, \ldots, N_{B_k}$ é a mesma que se obtém distribuindo n partículas independente e uniformemente sobre B. Este fato é muito útil na resolução de alguns problemas em que o processo de Poisson atua como entrada de um sistema mais complicado. Não entraremos em maiores detalhes deste aspecto do processo de Poisson no presente texto. (Veja entretanto os Exercícios 21 e 31 para ilustração de seus usos.)

9.4. DISTÂNCIA ÀS PARTÍCULAS

Suponha que temos um processo de Poisson sobre um subconjunto S do espaço Euclidiano. Se S tem volume finito, o número N de partículas em S é finito. Sejam $D_1 \leqslant D_2 \leqslant \cdots \leqslant D_N$ as distâncias da origem às partículas, ordenadas em ordem não-decrescente. Se S tem volume infinito, o número de partículas em S é infinito e representamos por $D_1 \leqslant D_2 \leqslant \cdots \leqslant D_m \cdots$ as distâncias da origem às partículas, ordenadas em ordem não-decrescente. Tal arranjo é possível porque, para qualquer número positivo r, apenas um número finito de partículas estão a uma distância menor que r da origem. Nesta seção obteremos a distribuição de D_m para diferentes escolhas do conjunto S.

Daremos inicialmente um exemplo em que essas distâncias ocorrem naturalmente: suponha que as estrelas situadas em um certo conjunto S do espaço Euclidiano tridimensional se distribuem de acordo com um processo de Poisson sobre S com parâmetro λ. Suponha além do mais que essas estrelas tem igual brilho. A quantidade de luz de uma estrela que chega à origem, é proporcional ao inverso do quadrado da distância da origem à estrela. Assim a quantidade de luz recebida de

uma estrela a uma distância r da origem é K/r^2 para alguma constante K. A quantidade de luz recebida da estrela mais próxima (e portanto aparentemente mais brilhante) é K/D_1^2. A quantidade total de luz é

$$\sum_m \frac{K}{D_m^2}.$$

Através de técnicas avançadas de probabilidade pode-se mostrar que se S é todo o espaço tridimensional, a soma da série acima é infinita com probabilidade 1. Este fato tem interessantes implicações em cosmologia.

Determinaremos agora a distribuição de D_m supondo por simplicidade que S tem volume infinito. Seja $S_r = S \cap x: |x| \leqslant r$ (isto é, seja S_r o conjunto de pontos em S que estão a uma distância não superior a r da origem), e seja $\varphi(r)$ o volume de S_r. O número N_{S_r} de partículas em S_r tem uma distribuição de Poisson de parâmetro $\lambda\varphi(r)$. O evento $\{D_m \leqslant r\}$ é mesmo o que o evento $\{N_{S_r} \geqslant m\}$. Assim pelas Equações (39) e (40) do Capítulo 5

$$P(D_m \leq r) = P(N_{S_r} \geq m) = \int_0^{\lambda\varphi(r)} \frac{t^{m-1}e^{-t}}{(m-1)!}\,dt. \tag{33}$$

Segue-se de (33) que se $\varphi(r)$ é diferenciável, então D_m tem função de densidade f_m dada por

$$f_m(r) = \frac{\lambda^m \varphi(r)^{m-1}\varphi'(r)e^{-\lambda\varphi(r)}}{(m-1)!}, \quad r > 0. \tag{34}$$

Se $\varphi(r)$ é estritamente crescente, tem uma função inversa contínua $\varphi^{-1}(r)$. Segue-se de (33) que a variável aleatória $\varphi(D_m)$ tem a distribuição gama $\Gamma(m, \lambda)$.

Em inúmeros casos importantes $\varphi(r)$ é da forma $\varphi(r) = cr^d$, onde c é uma constante numérica (isto é verdade, por exemplo, se $S = R^d$ ou se $S = [0, \infty)$). Neste caso (34) se transforma em

$$f_m(r) = \frac{d(c\lambda)^m}{(m-1)!} r^{md-1}e^{-c\lambda r^d}, \quad r > 0. \tag{35}$$

Diversos casos especiais desta fórmula serão considerados nos exercícios.

Usaremos a Fórmula (35) para determinar os momentos ED_m^j nestes casos. Assim

$$ED_m^j = \int_0^\infty r^j f_m(r)\,dr = \int_0^\infty \frac{d(c\lambda)^m}{(m-1)!} r^{md+j-1}e^{-c\lambda r^d}\,dr.$$

Para determinar estes momentos, fazemos a mudança de variável $s = \lambda c r^d$ e integramos para obter

$$ED_m^j = \frac{(\lambda c)^{-j/d}\Gamma(m + (j/d))}{(m - 1)!}.$$

9.5. TEMPOS DE ESPERA

Até este ponto visualizamos um processo de Poisson como um modelo para a distribuição de partículas no espaço. Se pensarmos no conjunto $[0, \infty)$ como sendo o eixo dos tempos, podemos considerar um processo de Poisson em $[0, \infty)$ como a distribuição de tempos em que ocorrem certos eventos. No começo da Seção 3, mencionamos alguns exemplos do uso de um processo de Poisson em $[0, \infty)$ desta maneira. Quando passamos a pensar em um processo de Poisson em $[0, \infty)$, como a distribuição de eventos no tempo e não mais como a distribuição de partículas no espaço, introduzimos um novo conjunto de termos. Em vez de falar de "partículas", passamos a falar de "eventos", e a distância D_m da origem à m-ésima partícula se transforma em tempo de ocorrência do m-ésimo evento.

Seja $N(t) = N_{[0, t]}$ o número de eventos que ocorrem no intervalo de tempo $[0, t]$. Então $N(t)$ é uma variável aleatória com distribuição de Poisson de parâmetro λt. Se $0 \leqslant s \leqslant t$, $N(t) - N(s)$ é o número de eventos que ocorre no intervalo de tempo $(s, t]$, e tem uma distribuição de Poisson de parâmetro $\lambda(t - s)$. De uma forma mais geral, se $0 \leqslant t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n$, então $N(t_1)$, $N(t_2) - N(t_1), \ldots, N(t_n) - N(t_{n-1})$ são variáveis aleatórias independentes que têm distribuições de Poisson de parâmetros

$$\lambda t_1, \lambda(t_2 - t_1), \ldots, \lambda(t_n - t_{n-1})$$

respectivamente. Estes fatos são decorrências imediatas da definição do processo de Poisson e sua descrição em linguagem de tempo.

Como foi mencionado acima, D_m é o tempo do m-ésimo evento. Sabemos dos resultados da Seção 9.4 que D_m tem distribuição gama $\Gamma(m, \lambda)$. Em particular D_1 se distribui exponencialmente com parâmetro λ. Lembre-se do Capítulo 6 que a soma de m variáveis aleatórias independentes identicamente distribuídas com distribuição exponencial têm a distribuição gama $\Gamma(m, \lambda)$. Suponha que se define as variáveis aleatórias $W_1, W_2, \ldots, W_n, \ldots$ da seguinte forma: $W_1 = D_1$, $W_n = D_n - D_{n-1}$, $n \geqslant 2$. Então claramente $D_m = W_1 + \cdots + W_m$. A discussão acima torna plausível que $W_1, W_2, \ldots, W_m$ sejam variáveis aleatórias independentes exponencialmente distribuídas com parâmetro comum λ. Este fato é verdadeiro, sendo uma propriedade bastante interessante e útil do processo de Poisson em $[0, \infty)$. Naturalmente a variável aleatória W_m não é nada mais

que o tempo entre o $(m - 1)$-ésimo e o m-ésimo eventos, de modo que os tempos $W_1, W_2, \ldots$ são *tempos de espera* entre eventos sucessivos de um processo de Poisson.

Teorema 1. Sejam $W_1, W_2, \ldots, W_n, \ldots$ os tempos de espera entre eventos sucessivos de um processo de Poisson em $[0, \infty)$ de parâmetro λ. Então W_n, $n \geq 1$, são variáveis aleatórias mutuamente independentes, exponencialmente distribuídas, como média comum λ^{-1}.

Demonstração. Seja f_n a densidade n-dimensional dada por

$$f_n(t_1, \ldots, t_n) = \begin{cases} \lambda^n e^{-\lambda t_n} & \text{para } 0 \leq t_1 \leq t_2 \leq \cdots \leq t_n, \\ 0, & \text{para outros valores de } t_1, \ldots, t_n. \end{cases}$$

Vemos do Exemplo 13 do Capítulo 6 que o teorema é verdadeiro se, e somente se, as variáveis aleatórias $D_1, \ldots, D_n$ têm densidade conjunta f_n. Isto é verdadeiro para $n = 1$ uma vez que D_1 se distribui exponencialmente com parâmetro λ.

Uma demonstração geral rigorosa é mais complicada. Antes de apresentá-la, daremos uma forma heurística de constatar este fato.

Seja $0 = t_0 < t_1 < \cdots < t_n$ e suponha que se escolhe $h > 0$ suficientemente pequeno para que $t_{i-1} + h \leqslant t_i$, $i = 1, 2, \ldots, n$. Então (ver Figura 4)

$$\begin{aligned}
(36) \qquad & P(t_i < D_i \leq t_i + h, 1 \leq i \leq n) \\
&= P(N(t_1) = 0, N(t_1 + h) - N(t_1) = 1, \ldots, \\
&\qquad N(t_n) - N(t_{n-1} + h) = 0, N(t_n + h) - N(t_n) \geq 1) \\
&= e^{-\lambda t_1}(\lambda h) e^{-\lambda h} \cdots e^{-\lambda(t_n - t_{n-1} - h)}[1 - e^{-\lambda h}] \\
&= \lambda^{n-1} h^{n-1} e^{-\lambda t_n}(1 - e^{-\lambda h}).
\end{aligned}$$

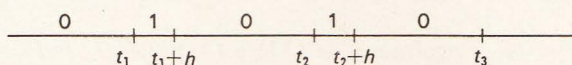



Figura 4

Se soubéssemos que as variáveis aleatórias $D_1, D_2, \ldots, D_n$ têm uma densidade conjunta g_n que é contínua no ponto $(t_1, t_2, \ldots, t_n)$, poderíamos concluir que

$$P(t_i < D_i \leqslant t_i + h \quad \text{para} \quad 1 \leqslant i \leqslant n) = g_n(t_1, \ldots, t_n) h^n + e(h),$$

onde $e(h)$ é alguma função de h tal que

$$\lim_{h \downarrow 0} \frac{e(h)}{h^n} = 0.$$

Seguir-se-ia então de (36) que

$$
\begin{aligned}
g(t_1, \ldots, t_n) &= \lim_{h \downarrow 0} h^{-n} P(t_i < D_i \le t_i + h) \\
&= \lim_{h \downarrow 0} \lambda^{n-1} e^{-\lambda t_n} \frac{1 - e^{-\lambda h}}{h} \\
&= \lambda^n e^{-\lambda t_n}
\end{aligned}
$$

como desejávamos.

Daremos agora uma demonstração elementar porém rigorosa do Teorema 1. Embora esta demonstração não seja difícil, é um tanto longa, e o leitor poderá omiti-la se assim o desejar.

Seja F_n a função de distribuição que tem densidade f_n. Da definição de f_n segue-se de imediato que para $n \geqslant 2$,

$$f_n(s_1, \ldots, s_n) = f_1(s_1) f_{n-1}(s_2 - s_1, \ldots, s_n - s_1).$$

Integrando ambos os membros sobre o conjunto $s_1 \leqslant t_1, \ldots, s_n \leqslant t_n$, vemos que

$$(37) \qquad F_n(t_1, \ldots, t_n) = \int_0^{t_1} f_1(s_1) F_{n-1}(t_2 - s_1, \ldots, t_n - s_1)\, ds_1.$$

Seja G_n a distribuição conjunta de $D_1, D_2, \ldots, D_n$. Do Exemplo 10 do Capítulo 6 vemos que o teorema é verdadeiro se, e somente se, as variáveis aleatórias $D_1, \ldots, D_n$ têm densidade conjunta f_n, e portanto a função de distribuição conjunta F_n. Conseqüentemente para demonstrar o teorema devemos mostrar que $F_n = G_n$.

Mas F_1 é simplesmente a distribuição exponencial de parâmetro λ. Como foi observado anteriormente, G_1, a distribuição de D_1, é também exponencial de parâmetro λ. Portanto $F_1 = G_1$.

Suponha que podemos mostrar que para $n \geqslant 2$,

$$(38) \qquad G_n(t_1, \ldots, t_n) = \int_0^{t_1} f_1(s_1) G_{n-1}(t_2 - s_1, \ldots, t_n - s_1)\, ds_1.$$

Então como $F_1 = G_1$, seguir-se-ia de (37) e (38) que $G_2 = F_2$. Usando o fato de que $G_2 = F_2$, outra aplicação de (37) e (38) mostraria que $G_3 = F_3$ etc. Em outras palavras, se soubéssemos que (38) é válido, o teorema seguir-se-ia de (37) e (38) por indução. Para estabelecer (38) podemos supor que $0 \leqslant t_1 \leqslant \cdots \leqslant t_n$, pois em caso contrário ambos os membros de (38) são zero.

Para iniciar a demonstração, observe que o evento $D_i \leqslant t_i$ é o mesmo que o evento $N(t_i) \geqslant i$ e assim

$$
\begin{aligned}
\{D_i \le t_i, 1 \le i \le n\} &= \bigcap_{i=1}^{n} \{D_i \le t_i\} \\
&= \bigcap_{i=1}^{n} \{N(t_i) \ge i\} \\
&= \{N(t_i) \ge i, 1 \le i \le n\}.
\end{aligned}
$$

Podemos, portanto, escrever

$$G_n(t_1, t_2, \dots, t_n) = P(N(t_i) \geq i, 1 \leq i \leq n).$$

Conseqüentemente (38) é o mesmo que

(39) $$P(N(t_i) \geq i, 1 \leq i \leq n)$$
$$= \int_0^{t_1} f_1(s_1) P(N(t_i - s_1) \geq i - 1, 2 \leq i \leq n)\, ds_1.$$

Para estabelecer (39) observe inicialmente que para qualquer $k \geqslant 1$

(40) $$\frac{(\lambda t)^k}{k!} e^{-\lambda t} = \int_0^t \lambda e^{-\lambda s} \frac{[\lambda(t - s)]^{k-1}}{(k - 1)!} e^{-\lambda(t-s)}\, ds.$$

De fato,

$$\int_0^t \lambda e^{-\lambda s} \frac{[\lambda(t - s)]^{k-1}}{(k - 1)!} e^{-\lambda(t-s)}\, ds = \frac{e^{-\lambda t}\lambda^k}{(k - 1)!} \int_0^t (t - s)^{k-1}\, ds$$
$$= \frac{e^{-\lambda t}\lambda^k}{(k - 1)!} \int_0^t s^{k-1}\, ds = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Suponha que $0 \leqslant t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n$ e que $1 \leqslant k_1 \leqslant k_2 \leqslant \cdots \leqslant k_n$. Alegamos a seguir que

(41) $$P(N(t_1) = k_1, \dots, N(t_n) = k_n)$$
$$= \int_0^{t_1} \lambda e^{-\lambda s} P(N(t_1 - s) = k_1 - 1, \dots, N(t_n - s) = k_n - 1)\, ds.$$

Para verificar este fato, observe que de acordo com (40)

(42) $$P(N(t_1) = k_1, \dots, N(t_n) = k_n)$$
$$= P(N(t_1) = k_1) \prod_{j=2}^{n} P(N(t_j) - N(t_{j-1}) = k_j - k_{j-1})$$
$$= \int_0^{t_1} \lambda e^{-\lambda s} \frac{[\lambda(t_1 - s)]^{k_1 - 1} e^{-\lambda(t_1 - s)}\, ds}{(k_1 - 1)!}$$
$$\times \prod_{j=2}^{n} \frac{[\lambda(t_j - t_{j-1})]^{k_j - k_{j-1}} e^{-\lambda(t_j - t_{j-1})}}{(k_j - k_{j-1})!}.$$

Por outro lado

(43) $$\int_0^{t_1} \lambda e^{-\lambda s} P(N(t_1 - s) = k_1 - 1, \dots, N(t_n - s) = k_n - 1)\, ds$$
$$= \int_0^{t_1} \lambda e^{-\lambda s} P(N(t_1 - s) = k_1 - 1)$$
$$\times \prod_{j=2}^{n} P(N(t_j - s) - N(t_{j-1} - s) = k_j - k_{j-1})\, ds$$
$$= \int_0^{t_1} \lambda e^{-\lambda s} \frac{[\lambda(t_1 - s)]^{k_1 - 1} e^{-\lambda(t_1 - s)}}{(k_1 - 1)!}$$
$$\times \prod_{j=2}^{n} \frac{[\lambda(t_j - t_{j-1})]^{k_j - k_{j-1}} e^{-\lambda(t_j - t_{j-1})}}{(k_j - k_{j-1})!}\, ds.$$

Comparando o segundo membro de (42) com o de (43), vemos que (41) é verdadeiro. A igualdade desejada (39) segue-se então de (41) pela soma de ambos os membros de (41) sobre todos os valores de $k_1, \ldots, k_n$ tais que $k_1 \leqslant k_2 \leqslant \cdots \leqslant k_n$ e $k_1 \geqslant 1$, $k_2 \geqslant 2, \ldots, k_n \geqslant n$.

Exercícios

1. Seja S_n um caminho aleatório com $\mu = 0$ e $S_0 = x$. Suponha que

$$P(a - c \leqslant S_T \leqslant b + d) = 1,$$

onde $a < x < b$, $c \geqslant 0$ e $d \geqslant 0$.

(a) Mostre que

$$(a - c)P(S_T \leq a) + bP(S_T \geq b)$$
$$\leq x \leq aP(S_T \leq a) + (b + d)P(S_T \geq b).$$

(b) Mostre que

$$\frac{x - a}{b - a + d} \leq P(S_T \geq b) \leq \frac{x - a + c}{b - a + c}.$$

2. Um jogador faz uma série de apostas de \$1. Ele decide abandonar o jogo quando seu ganho líquido alcança \$25 ou sua perda líquida alcança \$50. Suponha que tanto a probabilidade de ganhar como a de perder uma aposta são iguais a 1/2.

   (a) Determine a probabilidade de que tenha perdido \$50 ao abandonar o jogo.
   (b) Determine sua perda esperada.
   (c) Determine o número de apostas que fará antes de abandonar o jogo.

3. Suponha que o jogador descrito no Exercício 2 está jogando roleta e que suas probabilidades reais de ganhar e de perder uma aposta são 9/19 e 10/19, respectivamente. Resolver (a), (b) e (c) do Exercício 2 usando as probabilidades reais.

4. Um jogador faz uma série de apostas com probabilidade p de ganhar e probabilidade $q > p$ de perder cada aposta. Ele decide jogar até ganhar M_1 dólares ou perder M_2 dólares, onde M_1 e M_2 são números inteiros positivos. Ele pode apostar 1 ou 1/2 dólar de cada vez. Mostre que é mais provável ganhar M_1 dólares antes de perder M_2 dólares apostando 1/2 dólar de cada vez. Que generalização deste resultado parece plausível?

5. Derive (14) resolvendo a apropriada equação de diferença finita.

6. Seja S_n um caminho aleatório simples com $p = q = 1/2$ e seja $a < b$. Obtenha $P_{\{a,b\}}(x, y)$ e $G_{\{a,b\}}(x, y)$ para $a < x < b$ e $a < y < b$.

7. Seja S_n um caminho aleatório simples com $p = q = 1/2$. Obtenha $P_{\{0\}}(x, y)$ e $G_{\{0\}}(x, y)$. para $x > 0$ e $y > 0$.
8. Seja S_n um caminho aleatório simples com $0 < q < p$. Obtenha $P_\varnothing(x, y)$ e $G_\varnothing(x, y)$ para $x > 0$ e $y > 0$.
9. Seja S_n um caminho aleatório simples com $0 < q < p$. Obtenha $P_{\{0\}}(-1, y)$ e $G_{\{0\}}(-1, y)$ para $y < 0$.
10. Suponha que pontos se distribuem no espaço tridimensional segundo um processo de Poisson de parâmetro $\lambda = 1$. Cada ponto do processo é tomado como o centro de uma esfera de raio r. Seja X o número de esferas que contém a origem. Mostre que X se distribui segundo uma distribuição de Poisson de média $(4/3)\pi r^3$.
11. Escolhe-se um ponto aleatoriamente em um círculo de raio R com centro na origem. Toma-se o ponto como centro de um círculo de raio X onde X é uma variável aleatória com densidade f. Determine a probabilidade p de que o círculo aleatório contenha a origem.
12. Suponha que se escolhe n pontos independente e uniformemente em um círculo de raio R com centro na origem. Toma-se cada ponto como o centro de um círculo cujo raio tem densidade f. Obtenha em termos de p do Exercício 11 a probabilidade de que exatamente k círculos contenham a origem.
13. Obtenha a resposta do Exercício 12 para o caso em que se substitui os n pontos por um número aleatório de pontos tendo uma distribuição de Poisson de média πR^2.
14. Suponha que se distribui aleatoriamente N bolas em r caixas, onde N tem uma distribuição de Poisson de média λ. Seja Y o número de caixas vazias. Mostre que Y se distribui binomialmente com parâmetros r e $p = e^{-\lambda/r}$. *Sugestão:* se X_i é o número de bolas na caixa, i, então $X_1, X_2, \ldots, X_r$ são variáveis aleatórias independentes com distribuição de Poisson de parâmetro λ/r.
15. Usando o resultado do Exercício 14 podemos derivar facilmente a probabilidade $p_k(r, n)$ de que exatamente k caixas estejam vazias quando se distribui n bolas aleatoriamente em r caixas. Para tanto observe que

$$P(Y = k) = \sum_{n=0}^{\infty} P(N = n)P(Y = k \mid N = n)$$

$$= \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} p_k(r, n)$$

e

$$P(Y = k) = \binom{r}{k} e^{-\lambda k/r}(1 - e^{-\lambda/r})^{r-k}.$$

Assim

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} p_k(r, n) = \binom{r}{k} e^{\lambda(r-k)/r}(1 - e^{-\lambda/r})^{r-k}.$$

Igualando os coeficientes de λ^n, derive novamente a Equação (16) do Capítulo 2.

16. Suponha que os tempos de falhas sucessivas de uma máquina forme um processo de Poisson em $[0, \infty)$ de parâmetro λ.

    (a) Qual é a probabilidade de que ocorra pelo menos uma falha no intervalo de tempo $(t, t+h)$, $h > 0$?

    (b) Qual é a probabilidade condicional de pelo menos uma falha até o tempo $t+h$, dado que nenhuma falha ocorreu até o tempo t?

17. Suponha que temos um processo de Poisson de parâmetro λ em $[0, \infty)$. Seja Z_t a distância de t ao ponto mais próximo à direita. Obtenha a função de distribuição de Z_t.

18. Suponha que temos um processo de Poisson de parâmetro λ em $[0, \infty)$. Seja Y_t a distância de t ao ponto mais próximo à esquerda. Faça $Y_t = t$ se não existe nenhum ponto à esquerda. Obtenha a função de distribuição de Y_t.

19. Para Z_t e Y_t dos Exercícios 17 e 18,

    (a) Mostre que Z_t e Y_t são independentes,

    (b) Obtenha a distribuição de $Z_t + Y_t$.

20. Suponha que partículas chegam a um contador de acordo com um processo de Poisson de parâmetro λ. Cada partícula provoca um pulso de duração unitária. O contador registra a partícula se, e somente se, não há nenhum pulso presente no momento da chegada. Obtenha a probabilidade de que uma partícula seja registrada entre os tempos t e $t+1$. Suponha que $t \geqslant 1$.

21. Considere um processo de Poisson de parâmetro λ em $[0, \infty)$ e seja T uma variável aleatória independente do processo. Suponha que T tem uma distribuição exponencial de parâmetro ν. Seja N_T o número de partículas no intervalo $[0, T]$. Obtenha a densidade discreta de N_t.

22. Resolva o Exercício 21 supondo que T tem distribuição uniforme em $[0, a]$, $a > 0$.

23. Considere dois processos independentes de Poisson em $[0, \infty)$ tendo parâmetros λ_1 e λ_2, respectivamente. Qual é a probabilidade de que o primeiro processo tenha um evento antes do segundo?

24. Suponha que n partículas se distribuem independente e uniformemente sobre um disco de raio r. Seja D_1 a distância do centro do disco à partícula mais próxima. Obtenha a densidade de D_1.

25. Determine os momentos da variável aleatória D_1 do Exercício 24. *Sugestão:* obtenha uma integral Beta por meio de uma mudança de variável.

26. Considere um processo de Poisson de parâmetro λ em R^r. Para um conjunto A que tem volume finito, seja N_A o número de partículas em A.

    (a) Determine EN_A^2.

    (b) Se A e B são dois conjuntos com volumes finitos, determine $E(N_A N_B)$.

27. Sejam $A_1, A_2, \ldots, A_n$ n conjuntos disjuntos com volumes finitos, e de uma forma semelhante sejam $B_1, B_2, \ldots, B_n$ n conjuntos disjuntos com volumes finitos. Para números reais $\alpha_1, \ldots, \alpha_n$ e $\beta_1, \ldots, \beta_n$, seja

$$g(x) = \sum_{i=1}^{n} \beta_i 1_{B_i}(x). \quad \text{e} \quad f(x) = \sum_{i=1}^{n} \alpha_i 1_{A_i}(x)$$

Para um processo de Poisson de parâmetro λ mostre que

$$E\left(\sum_{i=1}^{n} \alpha_i N_{A_i}\right)\left(\sum_{i=1}^{n} \beta_i N_{B_i}\right)$$

$$= \lambda^2 \left(\int_{R^r} f(x)\,dx\right)\left(\int_{R^r} g(x)\,dx\right) + \lambda \int_{R^r} f(x)g(x)\,dx.$$

28. No Exercício 27 mostre que

$$\text{Var}\left(\sum_{i=1}^{n} \alpha_i N_{A_i}\right) = \lambda \int_{R^r} f^2(x)\,dx.$$

29. Considere um processo de Poisson de parâmetro λ em R^3 e seja D_m a distância da origem à m-ésima partícula mais próxima.

   (a) Obtenha a densidade de D_m.

   (b) Obtenha a densidade de D_m^3.

30. Suponha que temos um processo de Poisson de parâmetro λ no semiplano superior de R^2, isto é, o processo de Poisson está no subconjunto $S = (x,y): y > 0$ de R^2.

31. Considere o seguinte sistema. Os tempos de chegada de partículas no sistema constituem um processo de Poisson de parâmetro λ em $[0, \infty)$. Cada partícula permanece no sistema durante um certo tempo independente dos tempos de chegada das partículas e independente dos tempos de permanência de outras partículas. Suponha que os tempos de permanência das partículas se distribuam exponencialmente com parâmetro comum μ. Seja $M(t)$ o número de partículas presentes no sistema no tempo t. Determine a distribuição de $M(t)$ através das etapas seguintes.

   (a) Suponha que uma partícula chega, de acordo com uma distribuição uniforme em $[0, t]$ e, permanece durante um tempo exponencialmente distribuído com parâmetro μ. Obtenha a probabilidade p_t de que a partícula esteja no sistema no tempo t.

   (b) Usando o fato de que, dado $N(t) = n$, as partículas se distribuem independente e uniformemente em $[0, t]$, mostre que

$$P(M(t) = k \mid N(t) = n) = \binom{n}{k} p_t^k (1 - p_t)^{n-k}.$$

   (c) Mostre a seguir que $M(t)$ se distribui de acordo com uma distribuição de Poisson de parâmetro $\lambda t p_t$.

RESPOSTAS DOS EXERCÍCIOS

CAPÍTULO 1

2. 18/37. **3.** 1/2. **4.** 1/8.

5. (a) $e^{-10\lambda}$, (b) $1 - e^{-2\lambda} + e^{-3\lambda} - e^{-5\lambda}$.

6. 3/10. **7.** 1/2. **8.** 3/10.

9. 5/12. **10.** 5/8, 3/8. **11.** 4/5.

12. 1/2. **13.** (a) 1/2, (b) 1/2.

14. 2/5. **15.** 5/29. **16.** 10/19.

18. (a) 19/25, (b) 35/38. **19.** 0.

20. (a) $\dfrac{r(r-1)}{(r+b)(r+b-1)}$, (b) $\dfrac{rb}{(r+b)(r+b-1)}$,

(c) $\dfrac{br}{(r+b)(r+b-1)}$, (d) $\dfrac{b(b-1)}{(r+b)(r+b-1)}$.

21. (a) 98/153, (b) 55/153. **22.** (b) 13/25, (c) 21/25.

23. (b) 2/5, (c) 9/10. **25.** (a) 1/12, (b) 17/36, (c) 6/17.

26. (a) 1/4, (b) 2/5, (c) 1/2. **27.** 14/23.

28. 4/9. **29.** 2/13. **30.** 1/3.

31. (a) $(r+c)/(b+r+c)$, (b) $b/(b+r+c)$.

37. $1 - (4/5)^{51} \cdot 56/5$.

38. (a) $\sum_{k=0}^{3} \binom{10}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{10-k}$, (b) $1 - \left(\frac{3}{4}\right)^5$.

39. .9976. **40.** 2. **41.** 75/91.

42. $\dfrac{6 \cdot 12^2}{13^4 - 12^4}$. **44.** $1 - (1/10)^{12}$.

45. (a) $\binom{12}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{10}$, (b) $\sum_{k=0}^{2} \binom{12}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{12-k}$.

46. $1 - (11/4)(3/4)^7$.

47. $\dfrac{\sum_{k=2}^{12}\binom{12}{k}\left(\frac{9}{10}\right)^k\left(\frac{1}{10}\right)^{12-k}}{\sum_{k=1}^{12}\binom{12}{k}\left(\frac{9}{10}\right)^k\left(\frac{1}{10}\right)^{12-k}}$.

CAPÍTULO 2

1. 64.

2. (a) 2^n, (b) $2(2^n - 1)$.

3. $1/n$.

4. $(10)_6/10^6$.

5. $(n - 1)(r)_{n-1}/r^n$.

6. $\left[\binom{r-N}{N}\Big/\binom{r}{N}\right]^{r-1}$

7. $2(n - k - 1)/n(n - 1)$.

8. $n(n + 1)/2$.

9. (a) $(n - 2)/n(n - 1)$, (b) $(n - 2)/(n - 1)^2$.

10. (a) $\binom{n}{2}n!\,n^{-n}$, (b) $\binom{n}{2}(n-1)!/(n-1)^n$, (c) $1/n$.

11. $\binom{n}{j}(r - 1)^{n-j}/r^n$.

13. $\dfrac{\binom{r}{n-1}}{\binom{r+b}{n-1}}\,\dfrac{b}{r + b - n + 1} = \dfrac{\binom{r}{n-1}b}{n\binom{r+b}{n}}$.

14. (a) $4q$, (b) $4 \cdot 10q$, (c) $13 \cdot 48q$,
(d) $13 \cdot 12 \cdot 4 \cdot 6q$, (e) $4 \cdot \binom{13}{5} q$, (f) $10 \cdot 4^5 q$,
(g) $13 \cdot \binom{12}{2} 4^3 q$, (h) $\binom{13}{2}\binom{4}{2}\binom{4}{2} \cdot 11 \cdot 4q$, (i) $13\binom{12}{3}\binom{4}{2}4^3 q$.
Sendo $q = \binom{52}{5}^{-1}$.

15. $8\Big/\binom{10}{3}$.

16. $4 \cdot (48)_{n-1}/(52)_n$.

17. (a) $(r - k)_n/(r)_n$, (b) $\left(1 - \frac{k}{r}\right)^n$.

18. $\binom{r-k}{n-k}\Big/\binom{r}{n}$.

20. Expandir os termos.

21. $\binom{40}{10}\Big/\binom{50}{10}$.

22. $\binom{4}{3}\binom{48}{23}\Big/\binom{52}{26}$.

23. $\binom{26}{2}^2\Big/\binom{52}{4}$.

24. $\left[\binom{32}{5} - \binom{16}{5}\right]\Big/\left[\binom{52}{5} - \binom{36}{5}\right]$.

25. $\left[\binom{n}{5} - \binom{n-3}{5}\right]\Big/\binom{n}{5}$.

26. $\binom{95}{2}\Big/\binom{100}{2}$.

27. $\binom{n}{k}\binom{r-n}{n-k}\Big/\binom{r}{n}$.

CAPÍTULO 3

1. $f(x) = \begin{cases} 1/10, & x = 0, 1, \ldots, 9, \\ 0 & \text{para outros valores de } x. \end{cases}$

2. $P(X + r = n) = p^r \binom{-r}{n-r}(-1)^{n-r}(1-p)^{n-r}, \quad n = r, r+1, \ldots.$

3. (a) $P(X = k) = \dfrac{\binom{6}{k}\binom{4}{n-k}}{\binom{10}{n}}, \quad 0 \le k \le 6,$

(b) $P(X = k) = \binom{n}{k}\left(\dfrac{3}{5}\right)^k \left(\dfrac{2}{5}\right)^{n-k}, \quad 0 \le k \le n.$

4. $1/(2^{N+1} - 2)$.

5. (a) 0,3, (b) 0,3, (c) 0,55, (d) 2/9, (e) 5/11.

6. (a) $(1-p)^4$, (b) $(1-p)^4 - (1-p)^8 + (1-p)^{10}$,
(c) $(1-p)^3 - (1-p)^6 + (1-p)^7 - (1-p)^{11}$.

7. (a) 3/4, (b) 1/10, (c) 1/25, (d) 3/50.

8. $P(X = k) = (2k-1)/144, \quad 1 \le k \le 12.$

9. $P(X = k) = (k-1)\Big/\binom{12}{2}, \quad k = 2, 3, \ldots, 12.$

10. $P(Y = x) = \begin{cases} p(1-p)^x, & x = 0, 1, \ldots, M-1, \\ (1-p)^M, & x = M. \end{cases}$

11. (a) $P(X^2 = k^2) = p(1-p)^k, \quad k = 0, 1, 2, \ldots,$
(b) $P(X + 3 = k) = p(1-p)^{k-3}, \quad k = 3, 4, \ldots.$

12. (a) $P(Y \le y) = \binom{y}{n}\Big/\binom{r}{n}, \quad y = n\ n+1, \ldots, r,$

(b) $P(Z \ge z) = \binom{r+1-z}{n}\Big/\binom{r}{n}, \quad z = 1, 2, \ldots, r-n+1.$

13. (a) 2/3, (b) 2/9, (c) 13/27.

14. (a) $\dfrac{N+2}{2(N+1)}$, (b) $\dfrac{1}{N+1}$.

15. (a) $P(\min(X, Y) = z) = \dfrac{2(N-z)+1}{(N+1)^2}, \quad z = 0, \ldots, N,$

(b) $P(\max(X, Y) = z) = \dfrac{2z+1}{(N+1)^2}, \quad z = 0, \ldots, N,$

(c) $P(|Y - X| = 0) = \dfrac{1}{N+1}$,

$$P(|Y - X| = z) = \frac{2(N + 1 - z)}{(N+1)^2}, \quad z = 1, \ldots, N.$$

16. (a) $\dfrac{p_2}{p_1 + p_2 - p_1p_2}$, (b) $\dfrac{p_1p_2}{p_1 + p_2 - p_1p_2}$.

17. (a) densidade geométrica de parâmetro $p_1 + p_2 - p_1p_2$,

(b) $\dfrac{p_1p_2}{p_1 - p_2}\left[(1 - p_2)^{z+1} - (1 - p_1)^{z+1}\right]$, $z = 0, 1, 2, \ldots$

18. (a) $g(x)\sum_y h(y)$, (b) $h(y)\sum_x g(x)$.

20. 5/72.

21. (a) $\dfrac{(2r)!}{x_1! \ldots x_r!\, r^{2r}}$, onde x_i são números inteiros não-negativos cuja soma é $2r$.

(b) $\dfrac{(2r)!}{2^r r^{2r}}$.

22. (a) densidade binomial de parâmetros n e $p_1 + p_2$,

(b) $\dbinom{z}{y}\left(\dfrac{p_1}{p_1 + p_2}\right)^{z-y}\left(\dfrac{p_2}{p_1 + p_2}\right)^{y}$.

23. $(53/8)e^{-5/2}$.

24. $(17/2)e^{-3}$.

25. (a) $1 - (5/6)^6$, (b) 4.

26. $p^r(1 - p)^{x_r - r}$.

30. $\dfrac{\dbinom{x-1}{i-1}\dbinom{y-x-1}{j-i-1}\dbinom{n-y}{r-j}}{\dbinom{n}{r}}$.

31. $P(X + Y = z) = \begin{cases} \dfrac{z-1}{N^2}, & 2 \le z \le N, \\ \dfrac{2N + 1 - z}{N^2}, & N + 1 \le z \le 2N, \\ 0 & \text{para outros valores de } z. \end{cases}$

32. $\Phi_X(t) = \dfrac{1}{N+1}\left(\dfrac{1 - t^{N+1}}{1 - t}\right)$, $t \neq 1$, e $\Phi_X(1) = 1$.

33. $f_X(x) = \begin{cases} \dfrac{\lambda^{x/2}e^{-\lambda}}{(x/2)!}, & x \text{ inteiro não-negativo par}, \\ 0 & \text{para outros valores de } x. \end{cases}$

35. $\dbinom{z}{y}\left(\dfrac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{z-y}\left(\dfrac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{y}$.

36. $\dfrac{(x + y + z)!}{x!\, y!\, z!}\left(\dfrac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3}\right)^{x}\left(\dfrac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_3}\right)^{y}\left(\dfrac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}\right)^{z}$.

37. (a) $e^{\lambda p(t-1)}$,

(b) densidade de Poisson de parâmetro λp.

CAPÍTULO 4

1. $(2N + 1)/3$.

2. $4p(1 - p)(1 - 2p)$.

3. $\lambda^{-1}(1 - e^{-\lambda})$.

4. 17.

6. $p^{-1}(1 - p)[1 - (1 - p)^M]$.

7. $M + p^{-1}(1 - p)^{M+1}$.

8. $EX = N/2$ e $\operatorname{Var} X = (N^2 + 2N)/12$.

10. 2.

14. $E(2X + 3Y) = 2EX + 3EY$,
$\operatorname{Var}(2X + 3Y) = 4 \operatorname{Var} X + 9 \operatorname{Var} Y$.

16. (a) $\left(1 - \frac{1}{r}\right)^n$, (b) $\left(1 - \frac{2}{r}\right)^n$, (c) $r\left(1 - \frac{1}{r}\right)^n$,

(d) $r\left(1 - \frac{1}{r}\right)^n\left[1 - \left(1 - \frac{1}{r}\right)^n\right] + r(r-1)\left[\left(1 - \frac{2}{r}\right)^n - \left(1 - \frac{1}{r}\right)^{2n}\right]$.

17. (a) 1, (b) 1.

18. $\displaystyle\sum_{i=1}^{k-1} \frac{i}{r(1 - i/r)^2}$.

20. $\dfrac{-\sigma_2^2}{\sqrt{(\sigma_1^2 + \sigma_2^2)(\sigma_2^2 + \sigma_3^2)}}$.

21. $9 - 2\sqrt{2}$.

22. -1.

23. (a) $-1/3$, (b) $-1/2$.

25. (c) $EXY = n(n-1)\dfrac{r_1 r_2}{r(r-1)}$,

$\operatorname{Var} X = n\left(\frac{r_1}{r}\right)\left(1 - \frac{r_1}{r}\right)\frac{r - n}{r - 1}$, $\operatorname{Var} Y = n\left(\frac{r_2}{r}\right)\left(1 - \frac{r_2}{r}\right)\frac{r - n}{r - 1}$;

(d) $\dfrac{n(n-1)\dfrac{r_1 r_2}{r(r-1)} - n^2\dfrac{r_1 r_2}{r^2}}{n\left(\dfrac{r-n}{r-1}\right)\sqrt{\dfrac{r_1 r_2}{r^2}\left(1 - \dfrac{r_1}{r}\right)\left(1 - \dfrac{r_2}{r}\right)}}$.

26. $\delta = 1$.

27. A desigualdade de Chebyshev mostra que $a = 781$ será suficiente (ver também a resposta do Exercício 46 do Capítulo 7).

32. $z/2$.

33. $z\lambda_2/(\lambda_1 + \lambda_2)$.

CAPÍTULO 5

1. $F_X(-1) + 1 - F_X(3)$.

2. $F(x) = 0, x < 0$; $F(x) = x/R^2, 0 \le x \le R^2$; e $F(x) = 1, x > R^2$.

3. $F(x) = 0, x < 0$; $F(x) = x^3/R^3, 0 \le x \le R$; e $F(x) = 1, x > R$.

4. $F(x) = 0, x < 0$; $F(x) = x/a, 0 \le x \le a$; e $F(x) =$ para $x > a$.

5. $F(x) = 0, x < 0$; $F(x) = (2hx - x^2)/h^2, 0 \le x \le h$; e $F(x) = 1, x > h$.

6. $F(x) = 0, x < s\sqrt{3}/2$; $F(x) = \sqrt{4x^2 - 3s^2}/s$, $s\sqrt{3}/2 \leq x \leq s$; e $F(x) = 1, x > s$.

7. $F(x) = 0, x < 0$; $F(x) = x^2/2, 0 \leq x < 1$; $F(x) = -1 + 2x - (1/2)x^2$, $1 \leq x \leq 2$; e $F(x) = 1, x > 2$.

8. $m = \lambda^{-1} \log_e 2$.

9. $t = -\log 0{,}9/100 \log 2 = 1{,}52 \times 10^{-3}$.

10. $F(x) = 0, x < 0$; $F(x) = x/a, 0 \leq x < a/2$; e $F(x) = 1, x \geq a/2$.

11. (a) 7/12, (b) 1/3, (c) 1/6, (d) 5/12, (e) 1/2.

12. (a) (iv) $F(x-) = F(x)$ para todo x;
(b) (ii) F é uma função não-crescente de x e (iii) $F(-\infty) = 1$ e $F(\infty) = 0$;
(c) (ii) F é uma função não-crescente de x, (iii) $F(-\infty) = 1$ e $F(\infty) = 0$, e (iv) $F(x-) = F(x)$ para todo x.

13. $F(x) = 0, x < -5$; $F(x) = (x + 10)/20, -5 \leq x < 5$; e $F(x) = 1, x \geq 5$.

14. $e^{-1} - e^{-2}$.

15. $f(x) = 1/2(|x| + 1)^2 = F'(x)$ para todo x.

16. $f(x) = 3x^2/R^3, 0 < x < R$; e $f(x) = 0$ para outros valores de x.

17. $f(x) = x, 0 < x < 1$; $f(x) = 2 - x, 1 < x < 2$; e $f(x) = 0$ para outros valores de x.

18. $f_Y(y) = f(y) + f(-y), y > 0$; $f_Y(y) = 0, y \leq 0$.

19. $f(x) = 2xg(x^2), x > 0$; e $g(y) = f(\sqrt{y})/2\sqrt{y}, y > 0$.

20. Se $\beta > 0$, então $f_Y(y) = \beta y^{\beta-1}$, $0 < y < 1$, e $f_Y(y) = 0$ para outros valores de x. Se $\beta < 0$, então $f_Y(y) = -\beta y^{\beta-1}$, $y > 1$, e $f_Y(y) = 0$ para outros valores de x.

21. $f_Y(y) = y^{-2} f((1 - y)/y), 0 < y < 1$, e $f_Y(y) = 0$ para outros valores de x.

23. $\varphi(x) = (x - a)/(b - a), -\infty < x < \infty$.

24. Y tem uma densidade exponencial de parâmetro λ/c.

25. Multiplicar g por 12.

26. $f_Y(y) = |b|/\pi(b^2 + (y - a)^2), -\infty < y < \infty$.

27. $F(x) = 0, x < -1$; $F(x) = 1/2 + 1/\pi \operatorname{arcsen} x, -1 \leq x \leq 1$; $F(x) = 1, x > 1$. $f(x) = 1/\pi\sqrt{1 - x^2}, -1 < x < 1$, e $f(x) = 0$ para outros valores de x

28. $f(x) = \lambda|x|e^{-\lambda x^2}, -\infty < x < \infty$.

29. $X - a$ e $a - X$ têm a mesma distribuição. $F(a - x) = 1 - F(a + x)$ para todo x.

30. $\Phi(x) = 1/2 + 1/2 \operatorname{erf}(x/\sqrt{2}), -\infty < x < \infty$.

31. $f_Y(y) = \dfrac{2}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2}$, $0 < y < \infty$, e $f_Y(y) = 0$ para outros valores de y.

32. $f_Y(y) = \dfrac{1}{\sigma y\sqrt{2\pi}} \exp[-(\log y - \mu)^2/2\sigma^2], 0 < y < \infty$, e $f_Y(y) = 0$ para outros valores de y.

33. 0,6826.

34. $(X - \mu)/\sigma$ tem a distribuição normal padrão.

35. $f_Y(-6) = 0{,}0030$, $f_Y(-5) = 0{,}0092$, $f_Y(-4) = 0{,}0279$, $f_Y(-3) = 0{,}0655$, $f_Y(-2) = 0{,}1210$, $f_Y(-1) = 0{,}1747$, $f_Y(0) = 0{,}1974$, $f_Y(1) = 0{,}1747$, $f_Y(2) = 0{,}1210$, $f_Y(3) = 0{,}0655$, $f_Y(4) = 0{,}0279$, $f_Y(5) = 0{,}0092$, $f_Y(6) = 0{,}0030$, $f_Y(y) = 0$ para outros valores de y.

36. $\mu = 160$. $\sigma = 29{,}6$ $P(X \geq 200) = 0{,}0885$, $P(X \geq 220 \mid X \geq 200) = 0{,}244$ (24.4%).

38. 2 segundos.

39. Com distribuição geométrica de parâmetro $1 - e^{-\lambda}$.

40. (e) $g(t) = \lambda$, $t > 0$, onde λ é o parâmetro da distribuição exponencial; (f) melhora para $\alpha < 1$, deteriora para $\alpha > 1$ e permanece o mesmo para $\alpha = 1$.

41. Densidade gama $\Gamma(\alpha, \lambda/c)$.

43. $f_Y(y) = \dfrac{2\lambda^\alpha}{\Gamma(\alpha)} y^{2\alpha-1} e^{-\lambda y^2}$, $y > 0$, e $f_Y(y) = 0$ para outros valores de y.

44. $\varphi(y) = \sqrt{y}$, $y \geq 0$.

45. $\varphi(x) = [\Phi^{-1}(x)]^2$, $-1 < x < 1$.

46. $\Phi^{-1}(0{,}1) = -1{,}282$, $\Phi^{-1}(0{,}2) = -0{,}842$, $\Phi^{-1}(0{,}3) = -0{,}524$, $\Phi^{-1}(0{,}4) = -0{,}253$, $\Phi^{-1}(0{,}5) = 0$, $\Phi^{-1}(0{,}6) = 0{,}253$, $\Phi^{-1}(0{,}7) = 0{,}524$, $\Phi^{-1}(0{,}8) = 0{,}842$, $\Phi^{-1}(0{,}9) = 1{,}282$.

47. $\mu + 0{,}675\sigma$.

48. 1.

49. 0,82.

CAPÍTULO 6

1. $F_{W,Z}(w, z) = F\left(\dfrac{w-a}{b}, \dfrac{z-c}{d}\right)$; $f_{W,Z}(w, z) = \dfrac{1}{bd} f\left(\dfrac{w-a}{b}, \dfrac{z-c}{d}\right)$.

2. $F_{W,Z}(w, z) = F(\sqrt{w}, \sqrt{z}) - F(-\sqrt{w}, \sqrt{z}) - F(\sqrt{w}, -\sqrt{z}) + F(-\sqrt{w}, -\sqrt{z})$ e $f_{W,Z}(w, z) = \dfrac{1}{4\sqrt{wz}}(f(\sqrt{w}, \sqrt{z}) + f(-\sqrt{w}, \sqrt{z}) + f(\sqrt{w}, -\sqrt{z}) + f(-\sqrt{w}, -\sqrt{z}))$ para $w, z > 0$ e $F_{W,Z}(w, z)$ e $f_{W,Z}(w, z)$ são iguais a zero para outros valores de w e z.

3. (a) 3/4, (b) 5/12, (c) 3/4; estes resultados são obtidos facilmente determinando as áreas dos quadrados unitários adequados.

4. $1 - e^{-1/2\sigma^2}$

5. 3/8.

6. 1/3.

7. X distribui-se exponencialmente com parâmetro λ. Y tem a densidade gama $\Gamma(2, \lambda)$.
$F_{X,Y}(x, y) = 1 - e^{-\lambda x} - \lambda x e^{-\lambda y}$, $0 \leq x \leq y$;
$F_{X,Y}(x, y) = 1 - e^{-\lambda y}(1 + \lambda y)$, $0 \leq y < x$; e $F_{X,Y}(x, y) = 0$ para outros valores de x e y.

8. (a) $\alpha > -1$, (b) $c = (\alpha + 1)(\alpha + 2)$,
(c) $f_X(x) = (\alpha + 2)(1 - x)^{\alpha+1}$, $0 < x \leq 1$, e $f_X(x) = 0$ para outros valores de x. $f_Y(y) = (\alpha + 2)y^{\alpha+1}$, $0 \leq y \leq 1$, e $f_Y(y) = 0$ para outros valores de y.

9. $c = \sqrt{15}/4\pi$. X distribui-se segundo $n(0{,}16/15)$ e Y distribui-se segundo $n(0{,}4/15)$.

10. $f_{Y-X}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z + x)dx.$

11. (a) $f_{X+Y}(z) = \dfrac{\lambda_1\lambda_2}{\lambda_1 - \lambda_2}(e^{-\lambda_2 z} - e^{-\lambda_1 z})$, $z > 0$, e $f_{X+Y}(z) = 0$, $z \le 0$.
(b) $f_{X+Y}(z) = 0$, $z \le 0$; $f_{X+Y}(z) = 1 - e^{-\lambda z}$, $0 \le z \le 1$;
$f_{X+Y}(z) = e^{-\lambda z}(e^{\lambda} - 1)$, $1 < z < \infty$.

12. $f_{X+Y}(z) = \dfrac{\alpha + 2}{2} z^{\alpha+1}$, $0 \le z \le 1$, $f_{X+Y}(z) = \dfrac{\alpha + 2}{2}(2 - z)^{\alpha+1}$, $1 < z \le 2$;
$f_{X+Y}(z) = 0$ para outros valores de z.

13. $f_{|Y-X|}(z) = \dfrac{2}{b - a}\left(1 - \dfrac{z}{b - a}\right)$, $0 < z \le b - a$, e $f_{|Y-X|}(z) = 0$ para outros valores de z.

14. $f_Z(z) = \dfrac{1}{|b|}\int_{-\infty}^{\infty} f\left(x, \dfrac{z - ax}{b}\right) dx$, $-\infty < z < \infty$.

15. $(\alpha_1 - 1)/(\alpha_1 + \alpha_2 - 2)$.

16. $n(\mu_2 - \mu_1, \sigma_1^2 + \sigma_2^2)$.

17. $f_R(r) = \dfrac{r}{\sigma^2} e^{-r^2/2\sigma^2}$, $r > 0$; e $f_R(r) = 0$, $r \le 0$.

18. $f_{XY}(z) = \int_{-\infty}^{\infty} \dfrac{1}{|x|} f(x, z/x)dx.$

20. $f_Z(z) = 2/\pi(1 + z^2)$, $z > 0$, e $f_Z(z) = 0$, $z \le 0$.

21. $f_{Y/X}(z) = 1/(1 + z)^2$, $z > 0$, e $f_{Y/X}(z) = 0$, $z \le 0$.

22. Densidade Beta de parâmetros α_1, e α_2.

23. (a) $f_{Y|X}(x) = \lambda e^{-\lambda(y-x)}$, $0 \le x \le y$, e $f_{Y|X}(y|x) = 0$ para outros valores de x e z.
(b) $f_{Y|X}(y \mid x) = (\alpha + 1)(y - x)^{\alpha}/(1 - x)^{\alpha+1}$, $0 \le x < y < 1$,
e $f_{Y|X}(y \mid x) = 0$ para outros valores de x e y.
(c) $f_{Y|X}(y \mid x) = n(y; x/8, 1/4)$.

24. $\Phi(3/2) = 0{,}933$.

26. Densidade Beta de parâmetros $\alpha_1 + y$ e $\alpha_2 + n - y$.

27. $f_Y(y) = \alpha\beta^{\alpha}/(y + \beta)^{\alpha+1}$, $y > 0$, e $f_Y(y) = 0$, $y \le 0$. A densidade condicional de Λ dado $Y = y$ é a densidade gama $\Gamma(\alpha + 1, \beta + y)$.

28. $f_Y(y) = \dfrac{\sqrt{2/\pi}}{\sigma^3} y^2 e^{-y^2/2\sigma^2}$, $y \ge 0$, e $f_Y(y) = 0$, $y < 0$.

30. $f_Y(y) = y^2/2$, $0 \le y < 1$; $f_Y(y) = -y^2 + 3y - 3/2$, $1 \le y < 2$;
$f_Y(y) = y/2 - 3y + 9/2$, $2 \le y \le 3$, e $f_Y(y) = 0$ para outros valores de y.
$P(X_1 + X_2 + X_3 \le 2) = 5/6$.

31. $f_{X_1,X_2,X_3}(x_1, x_2, x_3) = 1/x_1x_2$, $0 < x_3 < x_2 < x_1 < 1$, e igual a zero para outros. valores de x_1, x_2 e x_3.

$f_{X_3}(x) = (\log_e x)^2/2$, $0 < x < 1$, e igual a zero para outros valores de x e y;

32. (a) $f_{X_1, X_n}(x) = n(n-1)(y-x)^{n-2}$ $0 < x \leq y < 1$, e igual a zero para outros valores de x e y;
(b) $f_R(r) = n(n-1)(1-r)r^{n-2}$, $0 < r < 1$, e zero para outros valores de r.
(c) Densidade Beta de parâmetros k e $n-k+1$.

33. Exponencial de parâmetro $n\lambda$.

34. $x^{(n/2)-1}\, e^{-x/2}/2^{n/2}\, \Gamma(n/2)$, $x > 0$, e 0 para outros valores de x.

35. Beta de parâmetros $m/2$ e $n/2$.

36. $aX + bY$ e $bX - aY$ distribuem-se conjuntamente como duas variáveis aleatórias independentes, cada uma com a densidade normal $n(0, a^2 + b^2)$.

37. $f_{X,X+Y}(x, z) = f(x) f(z - x)$.

38. Uniforme em $(0, z)$ para $z > 0$.

39. Uniforme em $(0,z)$ para $0 < z \leqslant c$, e uniforme em $(z-c, c)$ para $c < z < 2c$.

40. (a) $n(0, 1)$,

(b) $$f_{U,Z}(u, z) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{u^2 - 2\rho uz + z^2}{2(1-\rho^2)}\right],$$

(c) $$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right)\right],$$

(d) $$n\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1),\ \sigma_2^2(1-\rho^2)\right).$$

41. $$f_{W,Z}(w, z) = \left(\frac{z}{w+1}\right)^2 f\left(\frac{z}{w+1}, \frac{wz}{w+1}\right).$$

CAPÍTULO 7

1. $\alpha_1/(\alpha_1 + \alpha_2)$.

2. Z terá expectância finita quando $\alpha_1 > 1$ e $\alpha_2 > 0$. Neste caso $EZ = \alpha_2/(\alpha_1 - 1)$.

3. $\sigma\sqrt{2/\pi}$.

4. $X_\varepsilon/$ tem uma distribuição geométrica de parâmetro $(1 - e^{-\lambda\varepsilon})$. $EX_\varepsilon = \varepsilon e^{-\lambda\varepsilon}/(1 - e^{-\lambda\varepsilon})$. $\lim_{\varepsilon\to 0} EX_\varepsilon = 1/\lambda$.

5. $EX^m = \Gamma(\alpha_1 + \alpha_2)\,\Gamma(\alpha_1 + m)/\Gamma(\alpha_1)\,\Gamma(\alpha_1 + \alpha_2 + m)$.
Var $X = \alpha_1\alpha_2/(\alpha_1 + \alpha_2 + 1)(\alpha_1 + \alpha_2)^2$.

6. $$\sqrt{2}\,\Gamma\left(\frac{n+1}{2}\right)\Big/\Gamma\left(\frac{n}{2}\right).$$

8. $\alpha_2(\alpha_1 + \alpha_2 - 1)/(\alpha_1 - 1)^2(\alpha_1 - 2)$ para $\alpha_1 > 2$.

9. $EY = 3/2\lambda$. Var $Y = 5/4\lambda^2$.

10. $EX = 2R/3$, $\operatorname{Var} X = R^2/18$.

11. $EX = 0$, $\operatorname{Var} X = R^2/4$.

12. $EZ = \sigma\sqrt{\pi/2}$, $\operatorname{Var} Z = \sigma^2(2 - \pi/2)$.

13. $EY = 2\sigma\sqrt{2/\pi}$, $\operatorname{Var} Y = \sigma^2(3 - 8/\pi)$.

14. $EX = 0$, $\operatorname{Var} X = 1/2$.

15. (a) $E\,|X| = \sigma\sqrt{2/\pi}$, $\operatorname{Var}|X| = \sigma^2(1 - 2/\pi)$;
(b) $EX^2 = \sigma^2$, $\operatorname{Var} X^2 = 2\sigma^4$;
(c) $Ee^{tX} = e^{\sigma^2 t^2/2}$, $\operatorname{Var} e^{tX} = e^{2\sigma^2 t^2} - e^{\sigma^2 t^2}$.

16. $Ee^{tX} = \left(\dfrac{\lambda}{\lambda - t}\right)^{\alpha}$ para $t < \lambda$.

17. $EX^r = \Gamma(\alpha + r)/\Gamma(\alpha)\lambda^r$ para $r > -\alpha$.

19. $EX_k = k/(n + 1)$, $\operatorname{Var} X_k = k(n - k + 1)/n + 1)^2(n + 2)$.

20. $ER = (n - 1)/(n + 1)$, $\operatorname{Var} R = 2(n - 1)/(n + 1)^2(n + 2)$.

21. $\rho = 1/4$.

22. $EZ = \mu\alpha/\lambda$, $\operatorname{Var} Z = \alpha(\sigma^2\alpha + \sigma^2 + \mu^2)/\lambda^2$.

25. $\rho_3 \geq 0{,}458$.

26. $E[Y \mid X = x] = x$, $0 < x < 1$; $E[Y \mid X = x] = 2 - x$, $1 \leq x \leq 2$; e
$E[Y \mid X = x] = 0$ para outros valores de x.

27. $E[X \mid Z = z] = \alpha_1 z/(\alpha_1 + \alpha_2)$ para $Z > 0$ é o caso contrário.

28. $E[\Pi \mid Y = y] = (\alpha_1 + y)/(\alpha_1 + \alpha_2 + n)$, $y = 0, 1, 2, \ldots, n$, e 0 para outros valores de y.

33. $P(X \leq x) \approx \Phi((\lambda x - \alpha)/\sqrt{\alpha})$.

34. (a) $EX_1^2 = \sigma^2$ e $\operatorname{Var} X_1^2 = 2\sigma^4$.
(b) $P(X_1^2 + \cdots + X_n^2 \leq x) \approx \Phi((x - n\sigma^2)/\sigma^2\sqrt{2n})$.

35. (a) 0,921. (b) 0,842. (c) 23,26. (d) 27,71.

36. 0,9773. **37.** 0,02. **38.** 0,0415. **39.** 0,0053.

40. (a) $f_X(x) \approx \lambda^{-1/2}\,\varphi((x - \lambda)/\sqrt{\lambda})$,
(b) $f_X(x) \approx \Phi((x + 1/2 - \lambda)/\sqrt{\lambda}) - \Phi((x - 1/2 - \lambda)/\sqrt{\lambda})$.

41. $1/\sqrt{n\pi}$.

42. $1/\sqrt{n\pi}$. A aproximação (15) não pode ser aplicada diretamente porque o máximo divisor comum do conjunto $x - 1/x$ é um valor possível de S_1 é dois e não um.

43. 0,133. **44.** 0,523. **45.** $n \approx 6700$. **46.** 551.

CAPÍTULO 8

1. $M_X(t) = (e^{bt} - e^{at})/(b - a)t$, $t \neq 0$, e $M_X(0) = 1$.

2. $e^{at}M_X(bt)$.

4. (a) $M_X(t) = [p/(1 - e^t(1 - p))]^{\alpha}$, $-\infty < t < \log(1/(1 - p))$.

5. (b) $(2n)!$

6. (a) $\dfrac{dM_X(t)}{dt} = npe^t(pe^t + 1 - p)^{n-1}$ e

$$\frac{d^2M_X(t)}{dt^2} = npe^t(pe^t + 1 - p)^{n-1} + n(n-1)p^2e^{2t}(pe^t + 1 - p)^{n-2}.$$

10. $e^{\lambda(e^{it}-1)}$.

11. $p/(1 - e^{it}(1 - p))$.

12. $[p/(1 - e^{it}(1 - p))]^n$.

13. $[\lambda/(\lambda - it)]^r$.

14. $\varphi_X(t) = \Phi_X(e^{it})$.

21. (a) $\varphi_{X+Y}(t) = e^{-2|t|}$ e $\varphi_{(X+Y)/2}(t) = e^{-|t|}$.

23. (b) $\lim_{\lambda\to\infty} P\left(\dfrac{X_\lambda - \lambda}{\sqrt{\lambda}} \le x\right) = \Phi(x), \quad -\infty < x < \infty.$

CAPÍTULO 9

2. (a) 1/3, (b) 0, (c) 1250.

3. (a) $((10/9)^{50} - (10/9)^{75})/(1 - (10/9)^{75}) \approx 0{,}93$,
(b) $\approx$ \$44,75,
(c) ≈ 850.

6. Para $x = y$

$$P_{\{a,b\}}(y, y) = 1 - \frac{b - a}{2(y - a)(b - y)}$$

e

$$G_{\{a,b\}}(y, y) = \frac{2(y - a)(b - y)}{b - a} - 1.$$

Para $x < y$

$$P_{\{a,b\}}(x, y) = \frac{x - a}{y - a}$$

e

$$G_{\{a,b\}}(x, y) = \frac{2(x - a)(b - y)}{b - a}.$$

Para $x > y$

$$P_{\{a,b\}}(x, y) = \frac{b - x}{b - y}$$

e

$$G_{\{a,b\}}(x, y) = \frac{2(y - a)(b - x)}{b - a}.$$

7. Para $x = y$
$P_{\{0\}}(y, y) = 1 - 1/2y$ e $G_{\{0\}}(y, y) = 2y - 1$.
Para $x < y$
$P_{\{0\}}(x, y) = x/y$ e $G_{\{0\}}(x, y) = 2x$.
Para $x > y$
$P_{\{0\}}(x, y) = 1$ e $G_{\{0\}}(x, y) = 2y$.

8. Para $x = y$

$$P_{\varnothing}(y, y) = 1 + q - p \quad \text{e} \quad G_{\varnothing}(y, y) = \frac{1 + q - p}{p - q}.$$

Para $x < y$

$$P_{\varnothing}(x, y) = 1 \quad \text{e} \quad G_{\varnothing}(x, y) = 1/(p - q).$$

Para $x > y$

$$P_{\varnothing}(x, y) = \left(\frac{q}{p}\right)^{x-y} \quad \text{e} \quad G_{\varnothing}(x, y) = \frac{(q/p)^{x-y}}{p - q}$$

9. $P_{\{0\}}(-1, -1) = q$ e $G_{\{0\}}(-1, -1) = \dfrac{q}{p}$.

Para $y < -1$

$$P_{\{0\}}(-1, y) = \frac{p - q}{q[(q/p)^y - 1]} \quad \text{e} \quad G_{\{0\}}(-1, y) = \frac{1}{q(q/p)^y}.$$

11. $p = \dfrac{2}{R^2} \displaystyle\int_0^{\infty} \left(\int_0^R xf(x + z)dx \right) dz.$

12. $\dbinom{n}{k} p^k(1 - p)^{n-k}.$

13. $\dfrac{(\pi R^2 p)^k}{k!} e^{-\pi R^2 p}.$

15. $p_k(r, n) = \dbinom{r}{k} \displaystyle\sum_{j=0}^{r-k} (-1)^j \binom{r-k}{j} \left(1 - \frac{j+k}{r}\right)^n.$

16. (a) $1 - e^{-\lambda h}$, (b) $1 - e^{-\lambda h}$.

17. $F_{Z_t}(x) = 0, x < 0;$ e $F_{Z_t}(x) = 1 - e^{-\lambda x}, x \geq 0.$

18. $F_{Y_t}(x) = 0, x < 0; F_{Y_t}(x) = 1 - e^{-\lambda x}, 0 \leq x < t;$ e $F_{Y_t}(x) = 1, x \geq t.$

19. (b) $F_{Y_t + Z_t}(x) = 0, x < 0; F_{Y_t + Z_t}(x) = 1 - e^{-\lambda x}(1 + \lambda x), 0 \leq x < t;$
e $F_{Y_t + Z_t}(x) = 1 - e^{-\lambda x}(1 + \lambda t), t \leq x < \infty.$

20. $\lambda e^{-\lambda}.$

21. $f_{N_T}(k) = v\lambda^k/(\lambda + v)^{k+1}, k = 0, 1, 2, \ldots$, e zero para outros valores de k.

22. $f_{N_T}(k) = \dfrac{1}{\lambda a} \left[1 - e^{-\lambda a} \displaystyle\sum_{j=0}^{k} \frac{(\lambda a)^j}{j!} \right], k = 0, 1, 2, \ldots$, e zero para outros valores de k.

23. $\lambda_1/(\lambda_1 + \lambda_2).$

24. $f_{D_1}(x) = \dfrac{2nx}{r^2} \left(1 - \dfrac{x^2}{r^2}\right)^{n-1}, \quad 0 \leq x \leq r$, e zero para outros valores de x.

25. $ED_1^m = r^m n!\, \Gamma\left(\dfrac{m}{2} + 1\right) \Big/ \Gamma\left(\dfrac{m}{2} + n + 1\right).$

26. (a) $\lambda^2|A|^2 + \lambda|A|$, (b) $\lambda^2|A|\,|B| + \lambda|A \cap B|$.

29. (a) $f_{D_m}(r) = 3(4\pi\lambda/3)^m r^{3m-1} e^{-4\pi\lambda r^3/3}/(m - 1)!, r > 0$, e zero para outros valores de r.
(b) Densidade gama $\Gamma(m, 4\pi\lambda/3)$.

30. (a) $f_{D_m}(r) = (\pi\lambda)^m r^{2m-1} e^{-\pi\lambda r^2/2}/2^{m-1}(m - 1)!, r > 0$, e zero para outros valores de r.
(b) $ED_m = \dfrac{(\lambda\pi/2)^{-1/2}\, \Gamma(m + 1/2)}{(m - 1)!}$ e $ED_m^2 = \dfrac{2m}{\pi\lambda}.$

31. (a) $p_t = \dfrac{1}{t} \displaystyle\int_0^t e^{-\mu(t-s)}\, ds = \frac{1 - e^{-\mu t}}{\mu t}.$

TABELA I

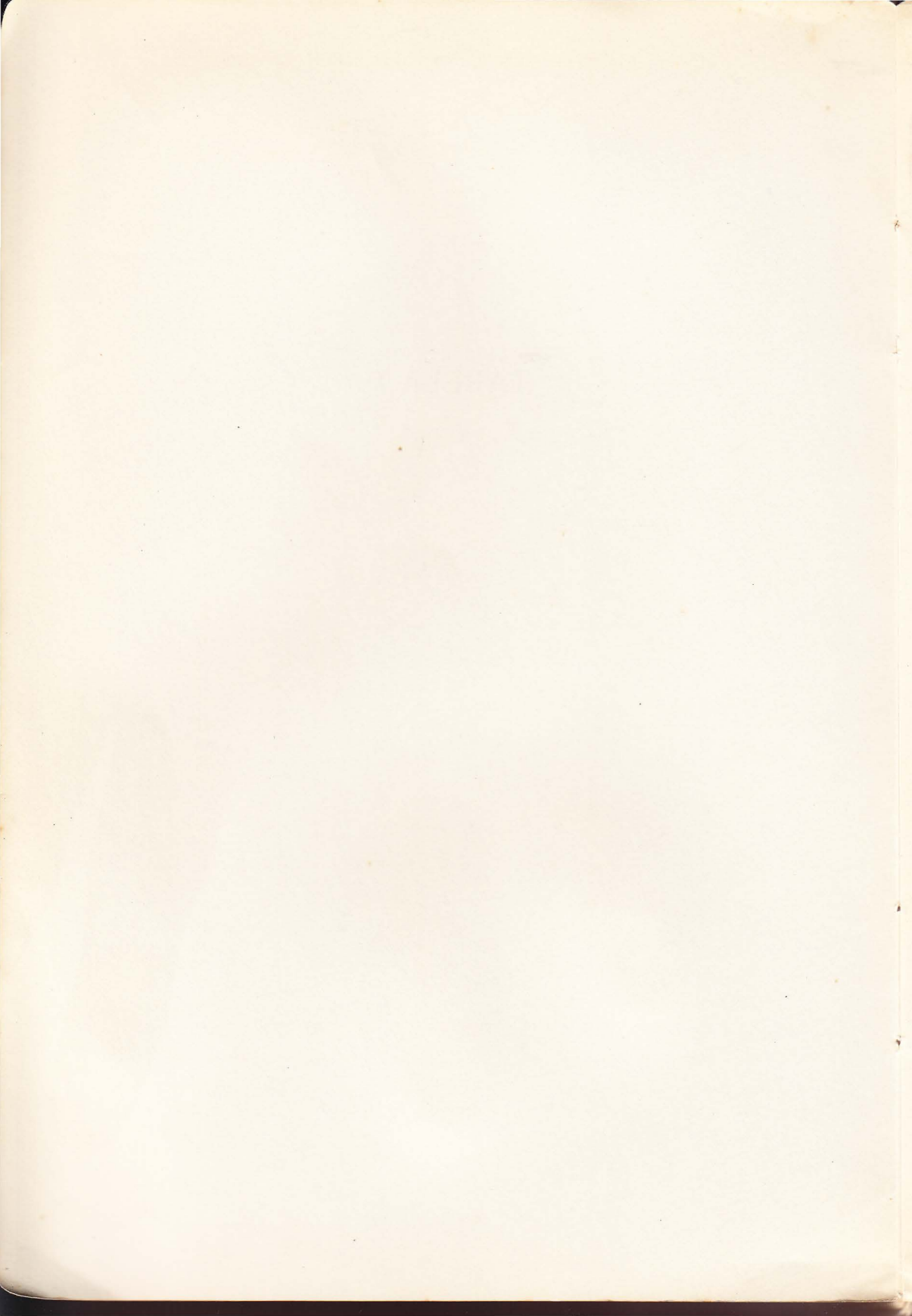



Tabela I Valores da função de distribuição normal padronizada

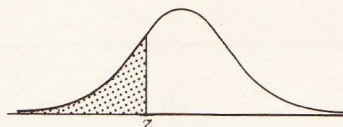
$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du = P(Z \leq z)$$

z	0	1	2	3	4	5	6	7	8	9
−3.	.0013	.0010	.0007	.0005	.0003	.0002	.0002	.0001	.0001	.0000
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0020	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0126	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0238	.0233
−1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0300	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0570	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
− .9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
− .8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
− .7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
− .6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
− .5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
− .4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3516	.3121
− .3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
− .2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
− .1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
− .0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Tabela I Valores da função de distribuição normal padronizada

z	0	1	2	3	4	5	6	7	8	9
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5363	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7974	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9430	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9648	.9656	.9664	.9671	.9678	.9686	.9693	.9700	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9762	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9874	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.	.9987	.9990	.9993	.9995	.9997	.9998	.9998	.9999	.9999	1.0000

Nota 1: Se uma variável normal X não está na forma padrão seus valores devem ser padronizados: $Z = (X - \mu)/\sigma$. Isto é $P(X \leqslant x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$.

Nota 2: Para probabilidades bicaudais ver Tabela 1b

Nota 3: Para $z \geqslant 4$, $\Phi(x) = 1$ para 4 casas decimais; para $z \leqslant -4$, $\Phi(z) = 0$ até 4 casas decimais.

Nota 4: As entradas opostas $z = 3$ são para 3,0; 3,1; 3,2; etc.

ÍNDICE REMISSIVO

E

F

i

j

k

l

m

n

p

q

r

s

t

u

v